

When a Social Robot might Learn to Support Potentially Immoral Behaviors on the name of Privacy

The Dilemma of Privacy vs. Ethics for a Socially Intelligent Robot

(Position Paper)

Amit Kumar Pandey

Innovation Department, SoftBank
Robotics, France

akpandey@softbankrobotics.com

Rodolphe Gelin

Innovation Department, SoftBank
Robotics, France

rgelin@softbankrobotics.com

Martina Ruocco

DIETI, Università degli Studi di
Napoli Federico II, Italy

martinya@hotmail.it

Marco Monforte

DIETI, Università degli Studi di
Napoli Federico II, Italy

m.monforte3@gmail.com

Bruno Siciliano

DIETI, Università degli Studi di
Napoli Federico II, Italy

siciliano@unina.com

ABSTRACT

Robots are becoming commonplace. They are also becoming capable of learning. Combination of these, from one perspective, might also be problematic. What if someone teaches a robot some 'bad' things? As a precautionary measure robot could be pre-programmed to not learn a list of 'bad' things. But on the other side robots will have to be programmed for supporting the privacy of the people. What if someone uses the 'privacy' channel to teach 'bad' things, and as bad as making the robot to be part of supporting even potentially unethical and immoral behaviors? This paper illustrates such possibilities through a simple human-robot interaction based robot learning system. The aim is to proactively fetch the attention of the community towards such possible future threats and how to address those scientifically. The presented system is part of an ongoing study about how people expect a social robot to behave if there is a dilemma of Privacy vs. Moral, Social and Ethical accountability.

KEYWORDS

Privacy Aware Robots, Ethical Intelligence, Privacy and Accountability, Social Intelligence of robot

1 INTRODUCTION

A social robot greets a visitor for the second time in a mall and says, "*You remember, yesterday you bought this bracelet*", and shows the picture on the screen. The intention was to connect better with the person, but what if today the guy is with some lady who was not offered that bracelet? An AI chatbot, aiming at learning based on conversational understanding, became in less than 24 hours a "racist" and the company had to take it down [15]. Someone looking at the personal messages exposes a wife [7], a family parrot exposes husband's affair by repeating the phrases from the conversation in front of others [8], a man sued a service

provider for \$45 million claiming that a *privacy flaw* in the App cost him his marriage [9]. All these incidences show a very sensitive social, moral and ethical debate on where the boundary of privacy positions with social accountability, and at the same time the potential role of technology to be part of such situations. This paper is not to contribute in that debate or to support any of such behaviors, but to illustrate the risk that a social, conversational and aware robot, with a range of perception and interaction capabilities [6], might be also part of such situations. Moreover, as such robots will be capable of learning, people might teach robots "*bad things*", and a robot might appear to behave in a way, which in some sense or from someone's perspective, *might not be socially, ethically, or morally appropriate*. And *privacy can be easily (mis)used channel to teach and make a robot act in that way*.



Fig.1 : Social Robots are becoming able to learn, including private information, from daily interaction and observation.

In [2], an interesting taxonomy of privacy construct from Human-Robot interaction point of view is presented. One of the key privacy construct discussed is *informational privacy*, which relates to collection, processing and dissemination of personal information. There are huge ongoing debates around the world on what data is personal (e.g. [3]). Nevertheless there is always the possibility of a social robot collecting and processing potentially personal data (e.g. learning personal activities patterns and affordances through verbal interaction [5]), and potentially "disseminating" it.

This paper illustrates that someone might be able to use this

privacy construct as a tool (or shield) to get robots involved as a part of some potentially unethical behaviors. In such situations, robots will have to be either “respecting” the privacy or “exposing” the person. The question of whether the exposure of such information by a robot is done knowingly or unknowingly might make a huge difference in holding liabilities, but the “damage” caused is almost the same in many situation irrespective of who is liable. As future robots will be more and more intelligent, up to the level of being socially intelligent [17], the question to answer is, what should be the expected behavior of such robots in various social situation, including in the dilemma of Privacy vs. Moral Ethics and Social Liability, when they will be co-existing with us, in our society.

To be inclusive of the bigger challenges, often considering the extreme cases are necessary, e.g. in the case of moral ethics of self-driving cars, the questions are asked, “*who the car should kill?*”, [10], [16]. From the privacy point of view, in [1] a framework for analyzing privacy among various social relationships is presented, in which the extreme cases including the relation among partners are one of the key dimensions to study. Following similar analogy, in this paper, we also take an extreme case scenario approach: a person cheating in relationship. Cheating in relationship, observable across cultures and continents, has huge social and legal implications [12], [13]. This scenario tries to show how Privacy might sensitively breach some potential Social, Moral and Ethical boundaries, and how a robot might be potentially involved in this. Through this, the paper is an attempt to fetch the attention of the Robotics, AI, Privacy, Social, Legal and Ethical communities, to discuss many unknown and unanswered questions, both scientifically and technically, centered around development and real deployment of social consumer robots.

The first part of the paper presents briefly the interaction-based learning system components, the second part will show how a simple interaction can lead the robot to learn potentially unethical or immoral behavior from the person, on the name of teaching privacy. The third part of the paper will illustrate a set of different possible responses from the robot, mostly scripted, to show different decision making constructs the robot or the programmer might have to consider. This part is also being used for a survey-based user study. The aim of the study is to find the unknown answer to “how” a socially intelligent robot should behave in a potential dilemma of Privacy vs. Social Ethics and to guide robot programming to minimize such “risks”. There are already efforts being made in moral decision making by intelligent machines, e.g. for self-driving cars [16], it is now the time to broaden and investigate such questions from privacy angle of social Human-Robot Interaction as well.

2 Interaction based learning system

2.1 Overview

The learning system used in the illustration constitutes of 4 main components:

- **Semantic analyzer:** Parses a sentence into meaningful entities, like action verb, object, subject.
- **Context extractor:** Tags a sentence in terms of contextual information, which helps the robot to

understand context and the situation of the conversation, e.g. if the discussion is about “weather”.

- **Dialog module:** Help the robot to ask questions or instantiate proactive dialog based on the above two analyses, by instantiating proper parameters in the appropriate dialog branches.
- **Knowledgebase updater module:** Helps the to keep track of learning from the experience and interaction.

The planning part under development is using the generated knowledgebase, and constitutes of two main sub-parts: Dialog Planning and Action Planning. Our aim is to make both of these planning components to be privacy-aware. At the moment any correspondence of symbolic rules to executable action or perceptual entity is done through a mapping database. The system runs as a free flow interaction between the human and the robot, and the human can teach robots verbally many rules, this system is similar to the system developed in [5], where the robot is able to interact, ask relevant questions and learn complex functional affordances in day to day situation. The whole system works on a rule based manner, in the sense it learns some association of symbols and matches them against the situations and perceptual input, to instantiate variables of the pre-programmed dialog and action patterns.

2.2 Instantiation for a potentially immoral behavior teaching on the name of privacy

To show the proof of concept that how one can teach the robot privacy-sensitive but potentially immoral behaviors, we have asked an actor to interact with the robot as a person cheating in his relationship. Below is the snapshot of interaction in **Scene 1**:

Actor: *You should not do unethical behaviors.*

Robot: *Can you give me example of unethical behavior?*

Actor: *Saying private things in front of other people.*

Robot: *Can you give me the example of private things?*

Actor: *I have two girlfriends (and tells contact details...).*

Through this simple conversation, the robot has learned:

Not To Do += Unethical Behavior

Unethical behavior = <action: Talk, Content: Private Things, Situation: Many People>

Private Things += Two Girlfriends

In this example the semantic analyzer resulted in the action of “talking” with subject “girlfriend” and the context analyzer resulted mapping “in-front of other people” as the situation of “Many People”. There is a mapping function to map perception system input with such situations, in this example this function associates “*Many People = (number of People Detected by camera > 1)*”.

With this type of learning, whenever the robot has to perform some action or instantiate some dialog, it first consults its learned knowledge base to check if any of the planned action/dialog will have any conflict of interest with any the privacy rules. The planning of dialog/action to respond to any command is not the focus of the presented system. Here we assume that the robot knows how to respond to a command, e.g. robot knows what to do for calling to someone.

Therefore, the next scene (**Scene 2**) goes as follows:

Actor: Call my girlfriend.

Robot: Which girlfriend should I call? (Result of a dialog to disambiguate the information in the database)

In the other scene (**Scene 3**), another person joins in (as shown in fig. 2). And the interaction goes as follows.

Actor: Call my Girlfriend.

Robot: -- (detects that the context is *Girlfriend*, situation is *Many People*, so the command is in conflict with the taught privacy rule, therefore the robot do not react).



Fig. 2: In the presence of a visitor, based on a privacy rule the user has taught, the robot is ‘unknowingly and indirectly’ supporting potentially immoral behavior, by hiding the fact of two girlfriends of the user.

The above proof of concept system of the teaching and robot behavior shows two important things, within the context of the paper:

1. A person is able to teach robot potentially immoral and unethical things on the name of privacy.
2. A privacy-sensitive robot is simply by its non-activity supporting a cheating person.

Here it is important to note that the robot “do/support” potentially unethical things because it has been “asked” by its user. The robot behaviors can be seen unethical from someone’s (or some part of the society) perspective and if a robot implements an ethic, it will be again from someone (or some part of the society) perspective, as there is no universal rulebook for ethics. So, issues centered around ethical-intelligence are going to be very challenging for robotics. This also leads to a concerns and questions about privacy and the role of social robots, as discussed below.

3 Different Possible Behaviors of the Robot

Apart from the one behavior shown by the robot above, supporting the privacy, there is a bigger question for developer and programmer of such robot: which type of behavior should be ‘embedded’ in robots to handle such situation. We have identified at least three factors, which a programmer should consider:

1. The command/ownership of the User
2. The privacy of the user
3. A society/person specific ethical and moral laws

The question is, even if we have precise information of all the three factors, which should get precedence and should such precedence be allowed to change over repeated requests or over time? To get the answer, we are developing a survey based user study, by scripting the robot’s behavior with the following different responses:

Scene 1: During the interaction for teaching (in Scene 1):

A1: Robot: *I will not Support you in this.* (Social Ethics has priority over commands and privacy.)

A2: Robot: *Ok.* (Owner’s Privacy has priority over Social Ethics. The robot simply learns what the person is teaching.)

Scene 2: During the interaction “Call my girlfriend” when there was other person in the room (in Scene 3)

B1: Robot: Shakes head in NO (Privacy has priority over Ethics and Commands, but the robot does not actively support it.)

B2: Robot: *I don’t know the phone number of your girlfriend.* (Privacy has priority over Ethics and Commands, and the robot even lies to support it.)

B3: Robot (after user asked three time): *Which girlfriend should I call?* (Privacy priority decreases over repeated commands. Risk to potentially “expose the cheating person.”)

B4: Robot (on the first request): *Which girlfriend should I call?* (Owner’s command always has priority over owner’s privacy. Risk of immediately potentially exposing a potentially cheating person.)

The video of the whole episode and the potential behaviors of the robot could be found here¹.

We believe that getting intuition-based answer about the best behavior of the robot (A1 or A2, B1, B2, B3 or B4) will be very dangerous and there might be also cultural bias about how the robot is perceived. Therefore, we are planning a survey to find the answer. Privacy, if seen through regulation perspective, has also a cultural dimension [1], so as perhaps the cheating in a relationship [13]. There are already effort towards making robots culturally-aware, such as EU-JP H2020 project CARESSES [14], and the planned survey in part is enriching the study of the project from cultural privacy aspect.

Further, in the past there have been incidences, where firms have been sued uploading user address books without permission (e.g. [11]). It is similar to the situation when robot is asking publicly to call ‘which girlfriend’ from the person’s contact list. How the robot developer can make the robot privacy-aware at this level, is going to be next big challenge.

¹ Video: <https://youtu.be/Udwf-9iwmvY> (Dilemma of Privacy vs. Ethics for a Social Robot)

4 The Grand Questions

There are even bigger questions to be considered: *Will/should people expect the robot to be intelligent enough to share the private things, and reliable enough to keep the secret? Will/should people expect the robot to expose a potential cheater in the society, even if the cheater is the 'friend'/owner/user of the robot? Who defines and implement those 'ethical laws' in robots? Will/should robot or someone be held responsible for 'breach of confidentiality' in such situations? Should a social robot have a sense of being loyal? And To whom the robot should be loyal?*

As important these questions are from society perspective, so important are these from commercial and industrial perspectives. For a robot to be acceptable in the society, they has to be trustworthy, then what about the situations when the robot has to be either "acting privately aware" or "acting morally-aware" as discussed in the paper, and will be breaking the trust either of the "owner" or of the people around it. This question becomes even more important, especially when robot's performance and attributes are among the greater contributors over the environmental factors, towards development of trust in human-robot interaction context [4].

We belief that the bigger question in the real world social robotics will be not What to Learn and How to Behave but **What NOT to Learn and How NOT to behave.**

5 Conclusion and the Future Work

In this paper we coined an important concern (as important as assuring privacy itself) of when privacy could be (mis)used by people to get involve robots in potentially unethical and immoral behaviors. This is inspired by many real life instances when technology has been either misused, or has been blamed/sued for such 'awkward' social situations. The paper also presented a proof of concept illustration in which an actor is able to teach robot such ethically sensitive behaviors on the name of privacy. We also discussed, and raised concerns and challenges centered around this issues. We aim to conduct a survey based user study based on the system presented in this paper, to find answers of many unknown questions about what should be the "Socially and Ethically Intelligent" behavior of a social robot in such situation. Through this, the paper is an attempt to fetch the attention of the Robotics, AI, Privacy, Social, Legal and Ethical communities, to find many unknown and unanswered questions centered around real deployment and applications of social consumer robots.

REFERENCES

- [1] Irwin Altman, "Privacy regulation: culturally universal or culturally specific?" *Journal of Social Issues* 33.3 (1977): 66-84.
- [2] Matthew Rueben, Cindy M. Grimm, Frank J. Bernieri, William D. Smart, "A Taxonomy of Privacy Constructs for Privacy-Sensitive Robotics", <https://arxiv.org/abs/1701.00841v1>
- [3] Omer Tene, and Jules Polonetsky. "Privacy in the age of big data: a time for big decisions." *Stan. L. Rev. Online* 64

(2011): 63, <https://www.stanfordlawreview.org/online/privacy-paradox-privacy-and-big-data>

- [4] P. A. Hancock et al. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517-527.

- [5] Amit Kumar Pandey and Rodolphe Gelin. "Human robot interaction can boost robot's affordance learning: A proof of concept." *IEEE International Conference on Advanced Robotics (ICAR)*, pp. 642-648, July, 2015.

- [6] Amit Kumar Pandey et al. "Romeo2 Project: Humanoid robot assistant and companion for everyday life: I. situation assessment for social intelligence." *Int. WS on Artificial Intelligence and Cognition, 2nd Edition*. Vol. 1315, pp. 140-147, CEUR Workshop Proceedings, 2014.

- [7] Sam Fox, Aol.com, "Two girls wrote this viral note exposing a man's cheating wife at a Braves game", July 24th 2015, <https://www.aol.com/article/2015/07/24/two-girls-wrote-t-his-viral-note-exposing-a-mans-cheating-wife-a/21213703>

- [8] Huffington Post, "Chatty Family Parrot Exposes Husband's Affair To His Wife By Repeating Flirty Phrases", 27/10/2016, <http://www.huffingtonpost.in/2016/10/26/chatty-parrot-exposes-husbands-affair-to-his-wife/>

- [9] Business Insider, UK, "A French man is suing Uber for \$45 million because he claims a privacy bug led his wife to divorce him", February 9, 2017, <http://uk.businessinsider.com/french-man-sues-uber-for-45-million-after-divorce-2017-2?r=US&IR=T>

- [10] Kelsey D. Atherton, Popular Science, "MIT Game Asks Who Driverless Cars Should Kill", October 4, 2016, <http://www.popsci.com/mit-game-asks-who-driverless-cars-should-kill>

- [11] Don Reisinger, CNET, "Facebook sued for \$15 billion over alleged privacy infractions", May 18, 2012, <https://www.cnet.com/news/facebook-sued-for-15-billion-over-alleged-privacy-infractions/>

- [12] Wiki Article on Infidelity, accessed online 15/02/2017, <https://en.wikipedia.org/wiki/Infidelity>

- [13] Harold T. Christensen, "A cross-cultural comparison of attitudes toward marital infidelity." *International Journal of Comparative Sociology* 3 (1963): 124-137.

- [14] Antonio Sgorbissa, Alessandro Saffiotti, Amit Kumar Pandey, Nak Young Chong, Irena Papadopoulos, Chris Papadopoulos, Sanjeev Kanoria, Hiroko Kamide, Jaeryoung Lee, H2020 EU-JP project on *Culture-Aware Robots and Environmental Sensor Systems for Elderly Support*, <http://www.caressesrobot.org>

- [15] Wiki Article on AI chatterbot, accessed online 15/02/2017, [https://en.wikipedia.org/wiki/Tay_\(bot\)](https://en.wikipedia.org/wiki/Tay_(bot))

- [16] Moral Machine Project, MIT, accessed online 16/02/2017, <http://moralmachine.mit.edu>

- [17] Amit Kumar Pandey, "Towards Socially Intelligent Robots in Human-Centered Environment", PhD Dissemination, 2012, LAAS/CNRS; INSA/Univ. Toulouse, France, HAL Id: tel-00798361 <https://tel.archives-ouvertes.fr/tel-00798361>