Respectfulness by Design

Dina Babushkina
Post-doctoral researcher in “Ethical Actions by Robots” & “Towards Responsible Artificial Intelligence"
University of Helsinki
Department of Political and Economic Studies
Unit of Practical Philosophy
dina.babushkina@helsinki.fi
Entry point


• Respect is a concept which helps to solve the problem of ethically acceptable behaviour for IoT devices
• Ethical concerns such as privacy, autonomy, transparency, and social order
• Four types of respect in the context of technology:
  o **Directive respect** = “adhering to rules and directives”, “respect as compliance, satisfying safety and regulatory requirements, or following preferences and commands expressed by users during the lifetime of the device”
  o **Obstacle respect** = giving up a goal when it interferes with preferences of the user (persuasive technology)
  o **Recognition respect** = system recognising that the user deserves a certain treatment because of his/her certain aspect (“We might think of this as including devices accommodating users’ religious or cultural values, such as helping them avoid using technology on the Sabbath [9], as well as requiring that devices refrain from treating users solely as means to data collection and advertising revenue (similar to one of the formulations of Kant’s categorical imperative, see sidebar)”.
  o **Care respect** = simulating love and concern for wellbeing. “Characterised as care respect, this also includes situations where someone or something takes an action that goes against the short term wishes or interests of someone else in order to promote their long term welfare… Examples include providing clear stopping points when users spend large amounts of time using services such as video streaming, or limiting how much money can be spent via microtransactions”.

+ New concept
+ Substantial claims
? Ethics
? Respect
Problems

• Expectations from social robots as players in interpersonal relationships: anthropomorphising & fine-tuning to the user’s emotional needs -> niche for artificial trustworthiness & respectfulness; need to rethink the very nature of such relationships and what is rational to expect of its parties

• Domain of respecting entities: the subject of respect has always been a person, that is an entity capable of cognitive attitudes (at least, a belief about respectworthiness of someone/something) and affective attitudes (emotive states produced by such belief). See e.g.:

> "While a very wide variety of things can be appropriate objects of one kind of respect or another, the subject of respect (the respecter) is always a person, that is, a conscious rational being capable of recognizing and acknowledging things, of self consciously and intentionally responding to them, of having and expressing values with regard to them, and of being accountable for disrespecting or failing to respect them (Dillon, 2018)."

Under this paradigm, robots cannot be subjects of respect.

• Relevance to ethics: Respect as such is not an ethical concept.

Questions

1. How can we talk about respect in its application to robots without the inverted commas? What does it mean for a robot to be respectful? How can we conceptualise respectfulness as an attribute of an animate artefact in such a way that we (a) won’t lose anything substantial from the phenomenon of respect itself, but at the same time (b) acknowledge the unique ontological status of “smart” robots?

2. Which kind of respectfulness are we ethically justified to expect from robots?
First approach

• Generally, respectfulness is seen as having two components: stance/attitude + action

• Robots (strong AI excluded) cannot respect (= exhibit an original attitude) but they can be respectful (= fit certain expectations)

• For a robot to be respectful, it means that its algorithms and hardware are designed by the principles of respect which subsequently manifest in its activities and interaction with its users. This means re-thinking the concept of respectfulness.

• The task is, then, to understand:

  o To what extent are having certain cognitive capacities (such as beliefs, acknowledgements, judgements, commitments) and emotive capacities (emotions and feelings) needed for being respectful? These will have to be reserved for people responsible for the crucial stages of the process that leads to the robot's functioning in an interpersonal relationship, such as the design and development of robots (e.g. programmers and engineers), its integration into society, and its professional use (e.g. medical staff or care-providers employing a robot in their practice).

  o Which part of being respectful can and should be played by robots, and in which way?
My suggestion

Robot respectfulness is a case of *mediated respectfulness* or *respectfulness by design*. By this I mean that a robot can only be attributed respectfulness in a proper sense if its interactions with persons reflect the respectful attitude of the humans involved in its design and operation.

Robot respectfulness consists of two elements:

- robotic actions (or other forms of operations) that are determined by principles that
- reflect the corresponding standing attitude of respect by humans who are involved in its design, implementation and professional use.

In my approach, a robot can be said to be respectful if *its functionality in the interactions with a human is determined by the considerations of respect for persons/humanity*
Reasons

- Being respectful is about acknowledging a value

- From ethical point of view (respect is not an ethical concept) we are looking for such value that can be universalized
We can dismiss

- Respect to authority (coercion, fear, slave paradigm)

- Respect to wishes or point of view (taking one’s point of view it into account in the decision-making process; this is too weak; more about preferences settings)

- Care (promoting well-being; not necessary connected)

- Admiration (aesthetic attitude; not necessary)
Values of personhood

A respectful robot must not undermine, at least, these attributes of a person:

- Intellect
- Rationality of emotional reactions and reactive attitudes
- Autonomy (when applicable)
- Personal integrity
- Trust in expertise
Respectulness as an attitude

How a robot, as a player in interpersonal relationships, will act must depend on the considerations of respect as the principles guiding the design, implementation, and use of the robot.

This stance—respect, as a standing attitude of humans responsible for the design, implementation, and professional use of robots, has three elements:

(a) recognition of,
(b) approval of and
(c) commitment (=preparedness to act)

to the veneration of the values of personhood.
Putting respect into robotic practice

How can a robot be respectful of persons in the sense that acknowledges and venerates these values? The answer is, through:

(a) its designer/implementor/professional user recognizing and being bound by the rules protecting these values and
(b) their sharing of this commitment with robots by provisioning the deterministic component in their algorithms that will ensure that robots act in a way that does not insult or degrade the humanity of its user.
The human components of robot respectfulness:

• Eliminate from the design of robots, from the conditions of their implementation and professional use, assumptions that violate the considerations of the respect for persons.

• Take such precautions that no deception or misinformation about capabilities, functionality, and the role of robots take place – so that technology does not create false expectations in consumers. This implies striving for transparency, explainability, and openness.

• Never count on users not being able to know or understand certain processes due to their not being experts in the field.

• Ensure that instead of helping people to overcome problems, the technology will not exploit people's weaknesses (such as emotional, e.g. loneliness, idiosyncratic sexuality, and intellectual, such as predisposition to rely on authoritative opinions).

The robotic components of robot respectfulness:

• The way a robot is presented to the users, the way it functions and is used by professionals must not inhibit a person's ability to reason and see facts, judge the truth, and evaluate the robot's decisions and choices, if she chooses to.

• When used as a substitute for a human in interpersonal relationships, the robot must not deceive the user about its ability to reciprocate, especially given the known tendency to anthropomorphize non-human beings.

• The way a robot is presented to the users, the way it functions and is used by professionals must not inhibit the autonomy in the individuals capable of it. This means, a person must have a right to refuse to be treated, aided, or otherwise interact with the robot and conditions must be met for her to be able to make an informed consent. By the latter I mean that she must have access to the description of the robot's functionality and its limitations. This is important for the user to able to form realistic expectations from the interaction with the robot. Furthermore, the robot must provide the possibility for its user to know parameters and considerations relevant to decisions that a robot makes on her behalf and be able to disagree or refuse treatment.