Social Robotics

Albert-Ludwigs-Universität Freiburg

Felix Lindner, Laura Wächter, Bernhard Nebel

SoSe 2019

Robot Ethics

Outline

- Can a robot be a moral agent?
- Should a robot be treated as a moral patient?
- Are robots capable of interactions?
- How are typical definitions of the term social robot formulated?

Moral Agents

Definition (Moral Agent)

A moral agent is a being who is capable of making moral judgments based on some notion of right and wrong, and who can be held responsible for its actions.

The following slides on responsibility are based on:
Gunkel, D. J.: Mind the Gap: Responsible Robotics and the Problem of Responsibility. Ethics and Information Technology, 2017.

Standard Approach: Instrumentalism

- Morality rests on human shoulders, and if machines changed the ease with which things were done, they did not change responsibility for doing them. People have always been the only moral agents. (Hall, 2001)
- Technology is a means to an end (Heidegger, 1977); it is not and does not have an end in its own right.
- The instrumentalist theory [...] is based on the common sense idea that technologies are tools. And tools are neutral.
- etc.

Instrumentalism: Summary

- Instrumentalism
  - Premise Computer systems, no matter how automatic, independent, or seemingly intelligent they may become, are not and can never be (autonomous, independent) moral agents.
  - Conclusion I It is logically incorrect to ascribe agency to something that is and remains a mere object under our control.
  - Conclusion II Holding a robotic mechanism or system culpable would not only be illogical but also irresponsible. (“It wasn’t me, it was the computer”)

Instrumentalism: Criticism

- Instrumentalism puts all technologies ranging from corkscrews to smartphones into one category.
- Its answers may be not satisfying when it comes to more complex machines, learning-based systems, and social robots.

Possible Attack: Wrong Category

- It is not true that “tools are simple machines and machines complex tools”
- A machine performs with its tools the same operations as the worker formerly did with similar tools (Marx, 1977).
- A machine is not an instrument to be used by a human, it is designed and implemented to take the place of the human.
- For instance, an autonomous car is not designed to replace a common car (the tool)—it is intended to replace the driver.

Possible Attack: Wrong attribution of responsibility

- AlphaGo beat an expert human player. Who won? Who actually beat Lee Serdol?
- Instrumentalism answers: The programmers of AlphaGo.
- But this sounds counter-intuitive, because the engineers who design and build learning-based systems have little idea what the system will eventually do once they are in operation.

Possible Attack: Missing category

- Instrumentalism implies a distinction between the who and the what.
- Thinking back to the Jibo clip:
  - What: “your house, your car, your toothbrush”
  - Who: “the things that really matter’ (family members)
  - “and something in between is this guy” (Jibo)
- Need for an ontological extension to include these “things in between”

Three ways of responding: Instrumentalism

- Instrumentalism 2.0: We still say all these new innovations are just tools and their creators take full responsibility.
  - Con: Slows down technological development.
  - Con: Things will be very different with social robots, like Jibo, that invite and are intentionally designed for emotional investment and attachment. Most likely, we do not want them to treat as mere tools. Case: US soldiers in Iraq and Afghanistan have formed surprisingly close personal bonds with their units’ Packbots, giving them names, awarding them battlefield promotions, risking their own lives to protect that of the robot, and even mourning their death.

Three ways of responding: Machine Ethics

- Machine ethics: We finally build machines that do respond to morally challenging situations, i.e., are responsible themselves.
- If it is the machine that is making the decision and taking action in the world with little or no direct human oversight, it would only make sense to hold it accountable (or at least partially accountable) for the actions it deploys and to design it with some form of constraint in order to control for possible bad outcomes.
  - Con: Even if a robot was fully equipped with all the rules from the Laws of War, and had, by some mysterious means, a way of making the same discriminations as humans make, it could not be ethical in the same way as is an ethical human. Ask any judge what they think about blindly following rules and laws. (Sharkey, 2012)

Three ways of responding: Hybrid

- Hybrid reponsibility: Distribute moral agency over both human and technological artifacts. (Hanson, 2009)
- Decisions are always made in networks of interactive elements, and those networks have always included other humans, organizations, institutions, etc. now also robots.
  - Con: Someone still has to decide which resposibilities are assigned to which elements of the network.

Outline

- Can a robot be a moral agent?
  - Not under instrumentalism, but two approaches could yield positive answers: machine ethics and hybrid responsibility.
- Should a robot be treated as a moral patient?
- Are robots capable of interactions?
- How are typical definitions of the term social robot formulated?

Moral Patients

Definition (Moral Patient)

A Moral Patient is a thing towards which moral agents can have moral responsibilities.

- Tradionally: All and only moral agents qualify as moral patients. Only those who can be held morally responsible for their actions deserve moral rights.
  - My duty to not harm you corresponds to my right to be not harmed by you.
- Challenged by Animal Rights Movement: Animals are moral patients without being moral agents. They deserve moral rights, because they have interests and can suffer (P. Singer).
- Can a similar argument be constructed for robot rights?

The following slides are based on:
Coeckelbergh, M.: Robot rights? Towards a social-relational justification of moral consideration. Ethics and Information Technology, 12(3):209–221, 2010.

Standard Arguments for Moral Consideration

- Deontological
  - Argument I: Giving rights to an entity implies that the entity in question has inherent worth and that therefore the entity needs to be treated as such irrespective of all other considerations.
  - Argument II: All entities that are experiencing 'subjects of a life' have rights.
- Utilitarian: All beings that are sentient and therefore are interested in not suffering deserve rights.

Possible Attack: Relevance

- As todays robots do not experience 'subjects of a life' nor are they sentient, these properties are irrelevant to the question how to think about giving moral considerations to existing robots.

Possible Attack: Marginal cases

- If particular properties are agreed upon as being sufficient for moral status and if not all humans share these properties (all the time), does that imply that these humans are not worthy of our moral concern (at that time)? E.g., consciousness, rationality, sentient, ...

Possible Attack: Determination and Epistemology

- Once we agree on properties sufficient for moral status, how can we know these are the correct properties? And how can we proof that a particular entity has these properties?

Indirect Argument

- Abusing robots is wrong not because it is a violation of rights or because on balance more suffering is created than with another act, but because we, as members of a moral community, do not exercise virtues such as compassion when abusing them.
- By abusing robots, we violate property right of other humans.
- Con: This solution may go against the intuition that the motivation for and justification of moral consideration should not have its source in our own well-being but at least also in the well-being of the object of moral consideration.

Social-Relational Argument

- Moral consideration is not based on ‘real’ intrinsic properties of an entity but it is attributed within social relations.
- Moral consideration is based on apparent features.
- Context dependent: Moral considerations greanted to entities in various concrete social relations and contexts.
- Subject dependent: Moral consideration resides in how the object appears to the subject.

Outline

- Can a robot be a moral agent?
  - Not under instrumentalism, but two approaches could yield positive answers: machine ethics and hybrid responsibility.
- Should a robot be treated as a moral patient?
  - Not under the standard ethical theories (deontology, utilitarianism), but under a virtue ethics or relational ethics treating robots as moral patients could be required.
- Are robots capable of interactions?
- How are typical definitions of the term social robot formulated?

Interactions

- Irrespective of the questions of moral agency and patiency, maybe all that counts is that we interact with robots.
- Which interactions are possible between robots and humans?

The following slides are based on:
Danaher, J., McArthur, N.: Robot Sex—Social and Ethical Implications. The MIT Press, 2017.

Interactions: Robot Sex

- What would it take for a sex robot to be a sex partner?
- In order to know this, we need to know what it is to have sex with a robot.
- In order to know this, we have to know what it is to have sex.

Interactions: Robot Sex

- Broadest sense: One can have sex with lots of things.
  - This is not very interesting. No new moral concerns would result from people having sex with e.g. vacuum cleaning robots or other robots just conceived of as sex toys.
- Narrower sense: Sex is what you have with all and only your sexual partners.

Interactions: Robot Sex

- First attempt: Two people have sex if and only if they have penile-vaginal intercourse.
  - Excludes sex with robots, but also:
  - Excludes homosexual sex.
  - Includes rape.

Interactions: Robot Sex

- Second attempt: Two people have sex if and only if they are involved in some distinctive exercise of shared sexual agency.
  - Sexual agency: One of two things is true: Either it pays the right sort of attention to sexual organs; or it involves a self-conscious understanding of the domain of the sexual whose boundaries may be idiosyncratic.
  - Shared agency: Doing something together with others, as opposed to alongside them. Example by Searle: Picknickers in the park running to shelter to avoid the rain may move in the same ways as a performance art troop.
- Implication: Having sex with a robot requires shared agency, i.e., a We that is involved in sexual activity. Particularly, having sex with a robot requires that you have sex with the robot and the robot has sex with you.

Other Interactions

- We could also have asked what it takes for a robot to have a conversation with a human or to go for a walk with a human.
- One can make the same argument by requiring shared going-for-a-walk agency, i.e., a We that goes for a walk rather than just two entities going for a walk in parallel.
- Keeping this in mind, let's look at typical definitions of the term social robot made up by social roboticists.

Outline

- Can a robot be a moral agent?
  - Not under instrumentalism, but two approaches could yield positive answers: machine ethics and hybrid responsibility.
- Should a robot be treated as a moral patient?
  - Not under the standard ethical theories (deontology, utilitarianism), but under a virtue ethics or relational ethics treating robots as moral patients could be required.
- Are robots capable of interactions?
  - In a broad sense, yes, but for genuine interactions shared agency must be established.
- How are typical definitions of the term social robot formulated?

Definitions: Robot Centered

Fong, Nourbakhsh, Dautenhahn (2003)

Social robots are embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other.

Bartneck, Forlizzi (2004)

A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact.

Criticism

- Social robots do not really fulfill the requirements formulated in the definitions.
- The conceptual norms that govern the semantics of the verbs highlighted—recognizing, engaging in social interactions, perceiving, interpreting, communicating, learning, following a norm —require that the subject of these verbs is aware, has intentionality or the capacity of symbolic representation, and understands what a norm is. Since robots—currently at least—do not possess such capacities—at least not how they are defined relative to our current conceptual norms—such characterizations are strictly speaking false. (Seibt, 2016)

Definitions: Human Centered

Breazeal (2003)

Augmenting such self-directed, creature-like behavior with the ability to communication with, cooperate with, and learn from people makes it almost impossible for one not to anthropomorphize them (i.e., attribute human or animal-like qualitities). We refer to this class of autonomous robots as social robots, i.e., those that people apply a social model to in order to interact with and to understand.

Breazeal (2002)

We interact with [a sociable robot] as if it were a person, and ultimately as a friend.

Breazeal (2002)

Ideally, people will treat Kismet as if it were a socially aware creature with thoughts, intents, desires, and feelings. Believability is the goal. Realism is not necessary.

Criticism

... the fictionalist interpretation of human-robot interactions collapses into what one might call the ‘error account’. Social robots are items that humans mistakenly engage in since a social interaction [...] requires the symmetric distribution of the capacity of understanding and following a norm. (Seibt, 2016)

... to treat something *as if* it were a person (a companion, a caregiver, a pet) is to take up the commitments that are attached to these social roles and treat it *as* a person (companion, caregiver, or pet). (Analogy: One cannnot fake a promise without actually making that promise.) (Seibt, 2016)

Outline

- Can a robot be a moral agent?
  - Not under instrumentalism, but two approaches could yield positive answers: machine ethics and hybrid responsibility.
- Should a robot be treated as a moral patient?
  - Not under the standard ethical theories (deontology, utilitarianism), but under a virtue ethics or relational ethics treating robots as moral patients could be required.
- Are robots capable of interactions?
  - In a broad sense, yes, but for genuine interactions shared agency must be established.
- How are typical definitions of the term social robot formulated?
  - Either, require robots to behave according to human norms (which may require too much from the robot)
  - Or, they only require robots to trigger human social behavior towards these robots (and may thus misprize what's really going on during interactions between robot and human)

Outlook

- In the remainder of this course, we will set the philosophical problems concerning social robots apart. Instead, we will take a look into how social robotics works as a field that is concerned with building robots and empirically investigating how humans respond to robots.