

P&M Robotics Debate 1st session - 23/07/2019

Is it possible to implement/program ethics in robotics?

Dynamics of the debate are as follows: a poll among the attendants is performed to know the a priori opinion. Each advocate has 3 minutes to present some of their points of view. Then, there is around 30 minutes of open debate, where the moderator will encourage particular questions of the public to the advocates. At the end, each advocate has 3 minutes for the closing remarks. Finally, a new poll will take place.

Advocates for NO	Alejandro Suárez and Irene García
Advocates for YES	Alberto Olivares and Pablo Jiménez
Moderator	Guillem Alenyà
Note taking	Júlia Borràs

Initial voting

YES	1
NO	The rest

Opening remarks

Pablo opening remarks (for YES)

- Ethics are already implemented implicitly.
- Explicitly, it seems clear it will be not possible to implement high level ethics concepts such as empathy, consciousness and emotions. But yes decision making.
- In a limited environment it is possible to implement some level of decision making that include some ethics.

Alejandro opening remarks (for NO)

- Decision making is not implementing ethics. Most ethical dilemmas are undecidable.
- Check the Moral Machine website where there are dilemmas that are unsolvable even for humans.

Alberto opening remarks (for YES)

- Emphasis on the feasibility: observe and extract rules and appropriate behaviors
- EU Guide: it identifies ethical issues on HRI, list the requirements of a robot to deal with such issues. It lists 7 principles
- Autonomy vs ethical compliance: Understand that ethical compliance that is very strict may restrict a lot the autonomy

Irene opening remarks (for NO)

- To generate a set of rules you need to consider all possible situations, which is impossible,
- Non-perfect ethics is not acceptable
- Ethics is an abstract concept with many parameters not even solvable by humans.
- Practically, the ethics of the robot are gonna be the ethics of the programmer, not the ones of all society.

Debate

Next, a list of short sentences with arguments that appeared during the debate:

- Behavior is more than rules.
- Ethics is evolving, robots can/should learn.
- Planning, a state machine of rules is too limiting, too many combinations.

- We can't just let robot learn, we need both the bottom up and the top down approaches.
- The rules written from the EU are suspiciously similar to Asimov rules. If the implementation is using First order logic rules, then is undecidable.
- Reason about what is good or bad is just general AI.
- Ethics implementation taking into account the context, that may change the result of what is good or bad, is already done in specific domains.
- You can't do a robot better than us. And you can't solve everything.
- If we imitate human behavior, is that already an ethical behavior? It is a simulation of a behavior.
- New decision criteria have to be learned? Yes, but must be supervised by humans to check if they are ethical.
- A common ground: ethics can be implemented up to some point. So, it is more a question on "to what extend" ethics can be implemented.
- Cost functions to evaluate situations.
- If it can't be done perfect, then what? We don't do it? Of course, yes we have to do it. But, are errors acceptable? (remember the self-driving cars example).
- We will accept killing machines because we already do.
- What about learning from the "most ethical people"?
- It seems that a solution would be bootstrap with expert knowledge and then learn from experience (with supervision).
- Teaching ethics to kids is implementing ethics.
- The bar is placed too high for ethics. We don't do that to everything. In our research, we try and our research is not approaching all the possibilities and it is definitely not perfect. Anything of what robots do is perfect, so, why Ethics need to be?
- We don't even know ourselves what is wrong or bad.
- Implementing a "too" ethical robot may lead to useless robots with too many ethical conflicts, not able to decide.

Closing remarks

Irene (for NO)

- Ethics is too global
- Capacity to decide if good or wrong is impossible
- A partial solution on ethical implementation is dangerous

Alberto (for YES)

- We've seen non-perfect solutions are feasible. Perfection is a utopia. Quote from Fernando
- Billy: "Utopia is ungraspable but it is useful to drive progress"
- Trying is worth. So, let's start talking towards this.

Alejandro (for NO)

- If you implement ethics with 1st order logic, you get undecidable problems
- Implementing ethics is equivalent to implementing AI

Pablo (for YES)

- A lot of work already done.
- Convinced that working with imperfect system and make them evolve, make them progress

Final vote

YES	7
NO	10

... and some undecided people.

Final remarks from Carme Torras

- We ask too much from machines
- We imagine robots facing extreme ethical dilemmas, but in contrast humans never really have to solve them
- Robots will have tunable rules, so that they can be adapted to each user.