

Ethics of Artificial Intelligence

How should we treat robots?

By Christine Randall

Background

- A robot can learn from its actions
- A robot cannot learn to change the way it was programmed—it is static, not dynamic
- Free will would be programmed in

Three Laws of Robotics

1. A robot cannot cause harm to a person through action or inaction (Moor)
2. Following a human's orders are necessary except when conflicting with first law
3. Protect itself as long as the first two laws are not conflicting

Introduction

- Background
- Reasons why robots can be shut down
- Reasons why robots should not be shut down
- Robots could affect our lives in a positive or negative way
- Evaluations

Background

- Humans think abstractly, programs have trouble.
- Humans learn through five senses, robots can have this information hard-coded into them.
- In determining how we treat robots, we must consider:
 - why we are making them and
 - if the way we make them entails that they should have their own rights
- Would robots be perfectly safe, or dangerous?
 - Military
 - Paper clips
- Robots must have unbreakable rules— three key laws of robotics

Questions To Consider

- Could robots “thinking” then override human’s moral views that were encoded into a robot?
- When should robots be responsible for their actions and judgments?
- Should artificial intelligence agents have legal rights or be persons in the eyes of the law?
- Why are we attempting to make ethical robots?
- If robots could think and feel, is it ethical to pull the plug?
 - Some say yes, we created them...
 - Others say no, they have a right to live...

Four Kinds of Ethical Robots (James Moor)

- ethical impact agents
 - Help or harm a human
 - Ex: a watch (on time)
- implicit ethical agents
 - security warnings and ethical behavior hard-coded into them
 - Ex: ATM, vending machine (“knows” stuff from memory chip, can’t answer “why”)
- explicit ethical agents
 - Come up with ethical decisions and work out conflicts
 - Ex: robots (“think for themselves,” can answer “why”)
- full ethical agents
 - reason a greater variety of topics
 - have consciousness, intentionality and free will
 - Far from implementing such a robot... (“think abstractly”)

Why robots can be shut down

- We wrote the code
- We built them to do work for us
- Robots do not have human level intelligence (Stefan Lovgren, co-creator of a female android)
- If you have an intelligent vehicle, who is responsible when it crashes?
- Maintenance reasons
- If we create a lifelike computer game character, it is still a game and it terminates whenever you feel like it
- Emotions in robots:
 - Code telling robots to be happy or sad
 - Illusion of emotions

Why robots should not be shut down

- Truly intelligent robots which would
 - Think and feel for themselves
 - Have feelings, wishes and morals
 - And even be able to reproduce (Charles Bailey)
- Some against since we would be playing God
- If we construct fully autonomous robots it would be more than dangerous
 - Would they be our allies or enemies?
 - They could destroy us on their own terms
- Emotions in robots:
 - Humans process according to rules, if rules bent, confusion persists (John Haugeland)
 - Emotions stored and based on experience
 - Robots need not have same emotions as humans (Nick Bostrom)

Pros and Cons

- Even if ethical, could potentially impact humans negatively
- Replace jobs
 - At first, negative
 - Later, positive
- Have wants of their own that is unwanted behavior (paper clip example)
- Could be destructive if fully autonomous

Kantian evaluation

- Rule: You can create an ethical autonomous robot which will gain knowledge for you
- It may come to realize that it shouldn’t have to work for you
- There are many potential dangers here
- Not ok to create the robots

Act utilitarianism

- Should autonomous ethical intelligent robots be created?
- Consequences—robots could be destructive instead of helpful
- Not ok to create the robots

Rule Utilitarianism

- Rule: You can create an autonomous ethical robot which will learn and gain knowledge for you
- The greatest good for the greatest amount of people would happen when the rule is not followed
- Robots would not always help humans like we created them for.
- Therefore the rule is not one to be followed.

Social Contract Theory

- If these autonomous intelligent robots were to kill humans, then the social order would not be maintained.
- Perhaps people would be afraid to go about their normal day.
- Therefore it autonomous robots should not be created.

Conclusion

- I believe we can create an ethical robot which does not have feelings, but simulates them
- If the robot did things only you wanted it to do and
 - did not do unethical things, and
 - did not think for itself but only help you (and search through its databases for you),
 - we could not really call it a person and it would be ok to create such a robot.
- I do not think we should try to create an autonomous robot which can “think for itself.”