The meaning of the EPSRC principles of robotics

Joanna J. Bryson (2017)

SAPIENZA
UNIVERSITÀ DI ROMA

Vincenzo Colella

The EPSRC principles of robotics

- In 2011, the **Engineering and Physical Sciences Research Council** (EPSRC) published a set of five ethical "**principles for designers, builders and users of robots**" in the real world.

- The best known set of laws are Isaac Asimov's "**Three Laws of Robotics**". These were introduced in 1942.

  i. *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
  ii. *A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.*
  iii. *A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.*

The principle of killing

I. *Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.*

- Robots have more than one use, as do knives.
- They should never be designed solely or principally to be used as weapons with deadly or other offensive capability.

BUT

- They are used as weapons from the government.

The principle of compliance

II. *Humans, not robots, are responsible agents. Robots should be designed and operated as far as is practicable to comply with existing laws, fundamental rights & freedoms, including privacy.*

- We have to make sure that robot actions are designed to obey the laws humans have made.
- A robot living in a human environment may be mistaken for a pet or other trusted family member, and therefore access private informations.
- It may have access to data that could unintentionally be stored in a public cloud, or a private cloud susceptible to hacking.

The principle of commoditization

III. Robots are products. They should be designed using processes which assure their safety and security

- Robots are pieces of technology that shouldn't be capable to defend themselves.
- Owners may certainly want to protect their robot, as they do with their car, house etc.
- The group chose to delete a phrase that had ensured the right of manufacturers or owners to build "self defence" capabilities into a robot.

The principle of transparency

IV. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.

- Robot intelligence is artificial, and humans must know that at all times.
- It is desirable for a robot to give the impression of real intelligence, but anyone who interacts with it should be aware of what it was really manufactured to do.
- One solution could be to print a code on the robot from which the owner is always able to look up for every functionality.

The principle of legal responsibility

V. The person with legal responsibility for a robot should be attributed.

- It should be possible to find out who is responsible for any robot.
- A robot is never legally responsible for anything.
- Every robot should have a searchable online licence which records the name of the manufacturer and the responsible human who acquired it.

Thoughts

- Who is Joanna J. Bryson ?
- Are the principles really useful ?
- What about Tesla’s crashes ?

The meaning of the EPSRC principles of robotics

Joanna J. Bryson (2017)

SAPIENZA
UNIVERSITÀ DI ROMA

Thank you !