ENARIS
Education and Awareness for Intelligent Systems

Scenario - Robot Butler

- It can do everything a human can do

  Help with chores, homework, gardening, ...
- You are its owner / master
- Reacts on its own following your commands
- **Needs rules on how to behave...**

… so it does not do anything wrong or harmful by accident

ENARIS
Education and Awareness for Intelligent Systems

Leading Questions

Should the robot be allowed to...

… **help** with your **homework**?

… **defend** against intruders?

… **activate itself** in case of an emergency?

… **leave** your home?

…

ENARIS
Education and Awareness for Intelligent Systems

Loopholes

- Do the rules prevent it from **doing your homework**? Even when you ask it step by step?
- **Who defines** what an **intruder** is? Could it be used to attack another home by redefining/faking ownership?
- Can it turn **immune to deactivation** in case of dangerous behaviour, based on the definition of an emergency? How is an emergency defined?