Robotic assistants:

AI, ethics & science fiction

Carme Torras

Institut de Robòtica i Informàtica Industrial

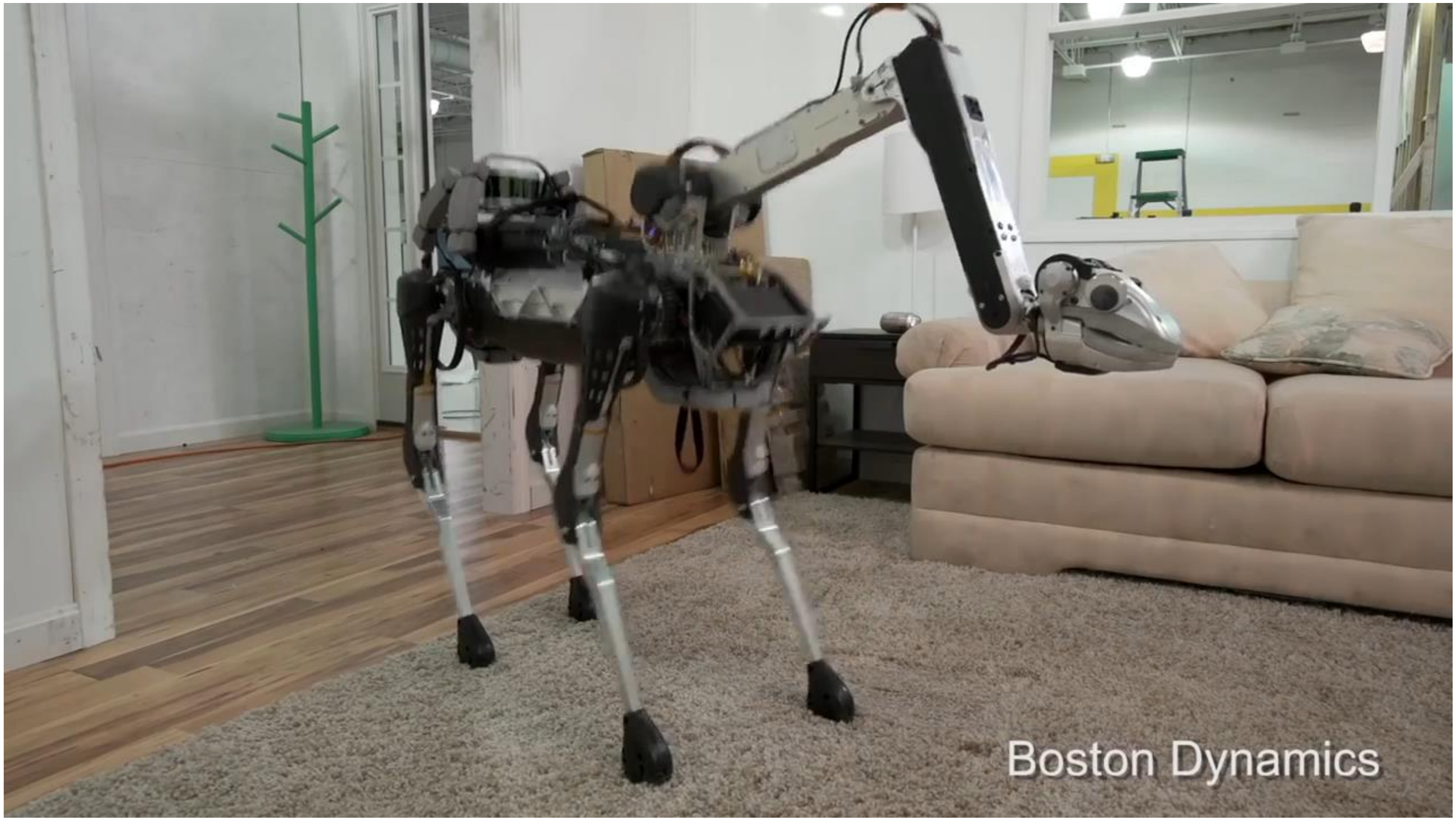
CSIC-UPC

Barcelona

Table of contents

Assistive robots

Domestic robot



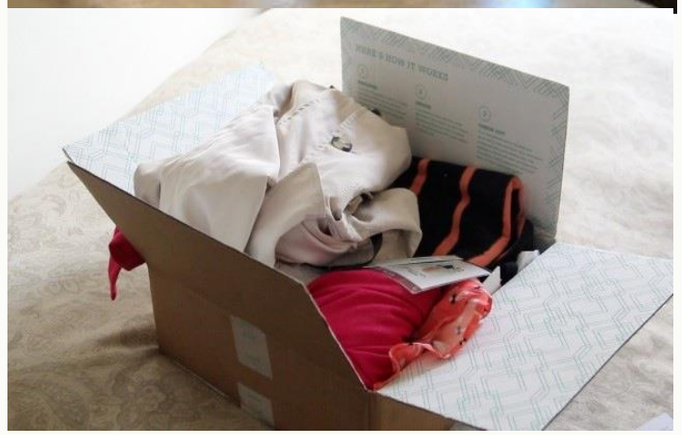
Domestic robot… and pet!

CLOTHILDE

CLOTH manIpulation Learning from DEmonstrations

housekeeping and hospital logistics

automation in the clothing and internet business



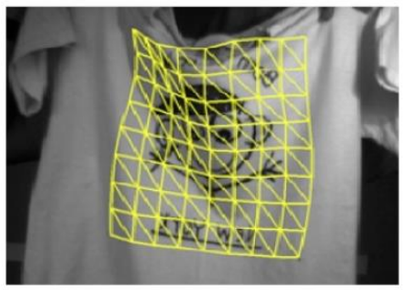
increasing the autonomy of the elderly and disabled

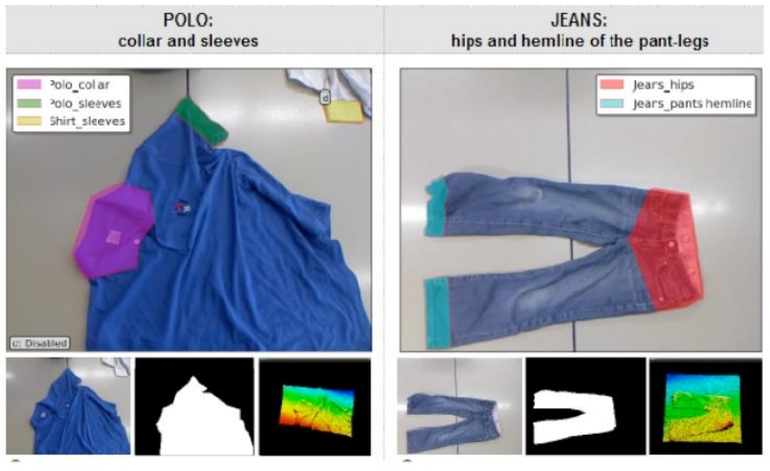
CLOTHILDE

CLOTH manIpulation Learning from DEmonstrations

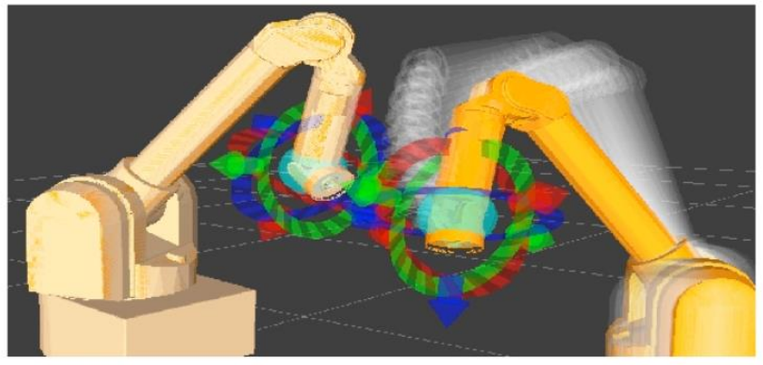
Theory of cloth manipulation based on computational topology

Cloth Perception





Manipulation and Planning



Learning from Demonstration

Research challenges in Robotic AI

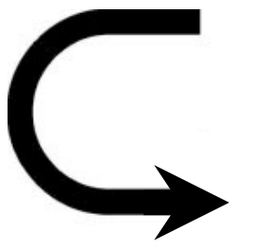
Industrial robots ⟶ Assistive robots

- Programmed by experts
- Caged
- Rigid objects
- Accurate
- Fixed sequences
- Non-interactive

- Easy to instruct by non-experts
- Intrinsically safe for people
- Able to perceive and manipulate non-rigid objects
- Tolerant to noisy perceptions and inaccurate actions
- Capable of goal-directed execution
- Collaborating with people

Usability

Easy instruction by non-experts



Exhaustive **programming** taking into account all situations

Learning from **demonstrations + reinforcement** learning

Usability

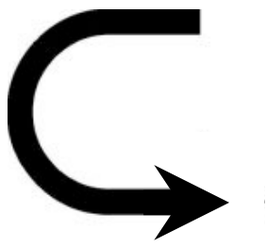
Easy instruction by non-experts

Exhaustive **programming** taking into account all situations

Learning from **demonstrations + reinforcement** learning

Usability

Easy instruction by non-experts



Exhaustive **programming** taking into account all situations

Learning from **demonstrations + reinforcement** learning

Personalization: **human in the** learning **loop**

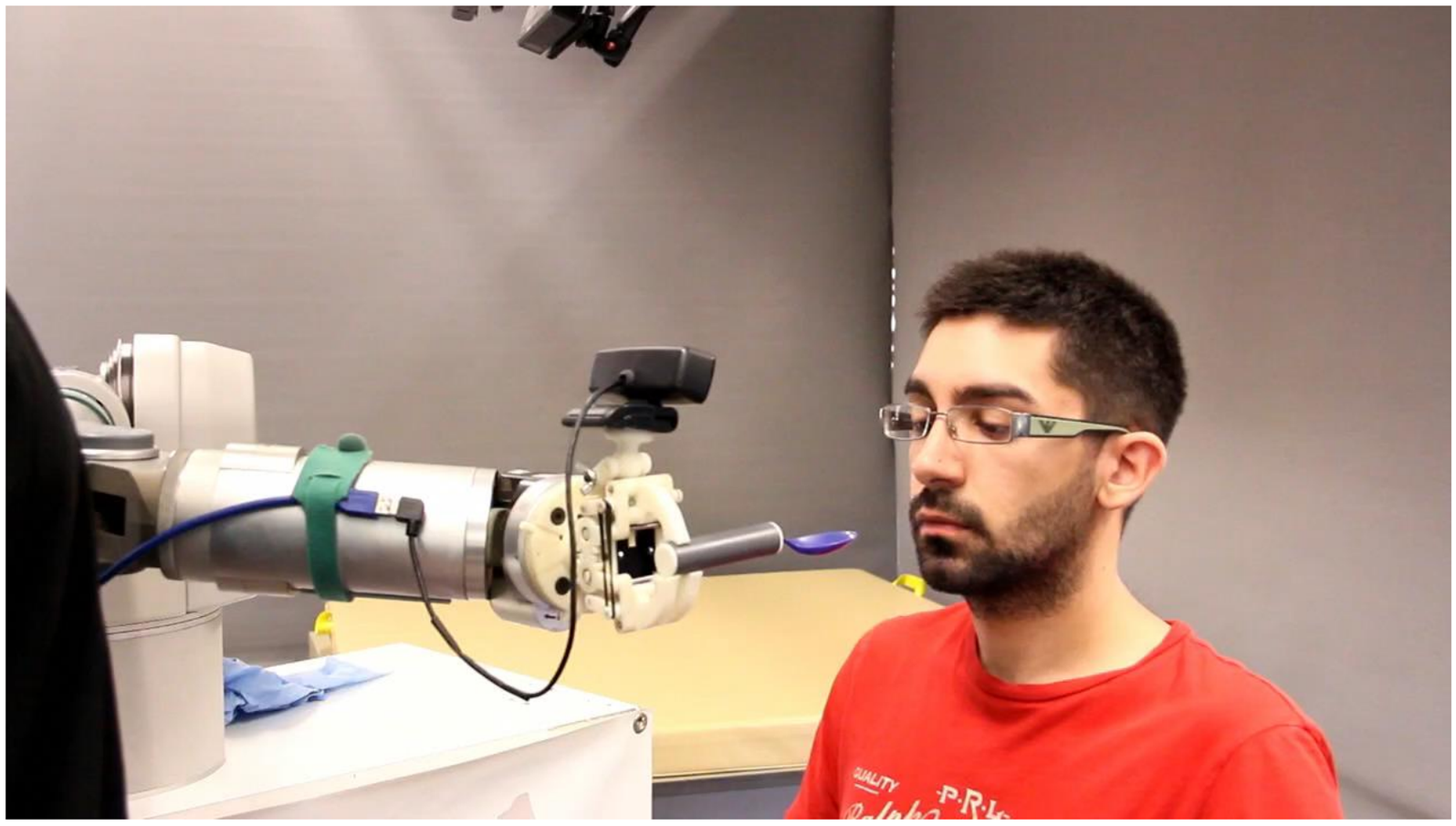
Usability

Easy instruction by non-experts

Exhaustive **programming** taking into account all situations

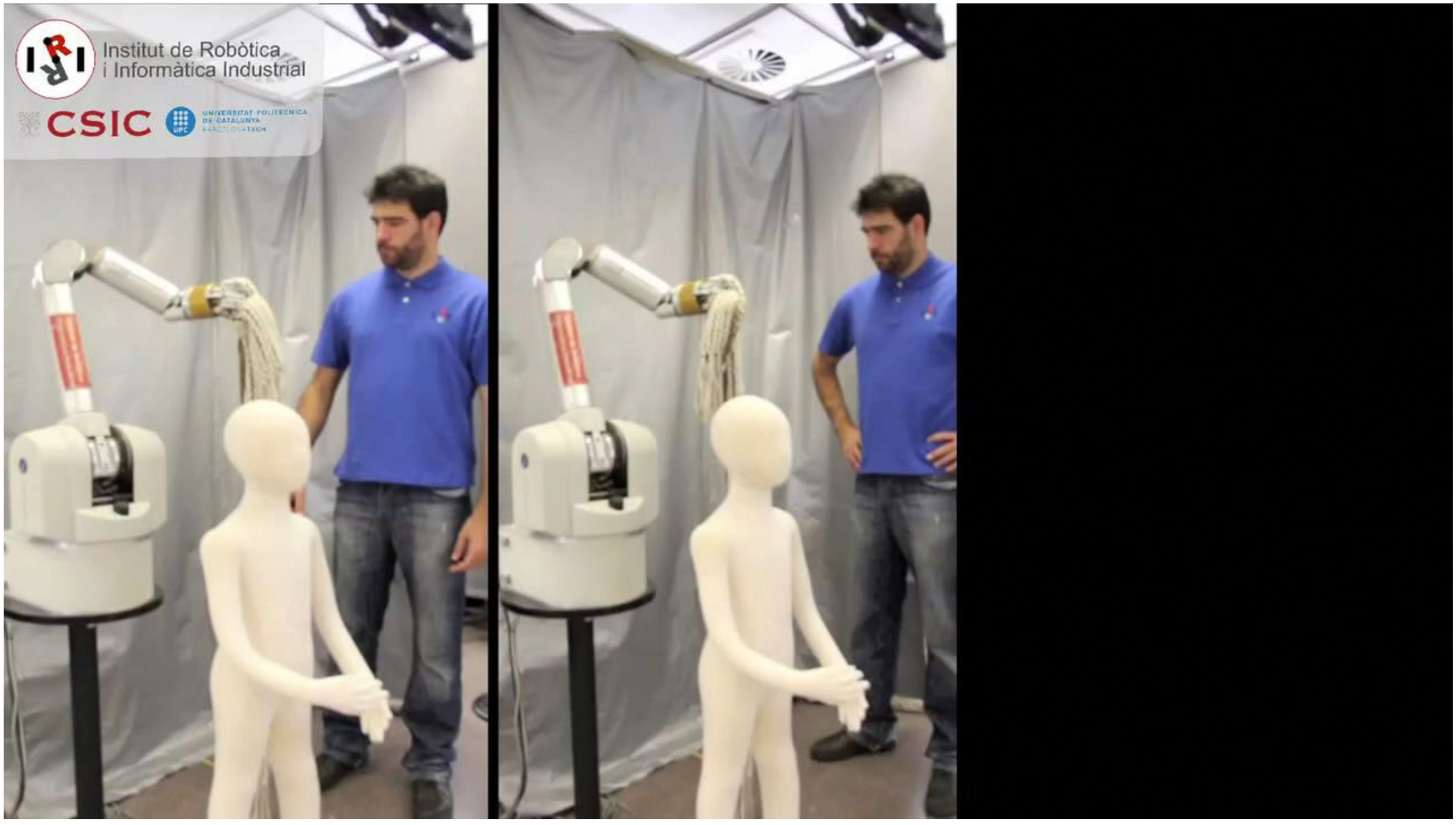
Learning from **demonstrations + reinforcement** learning

Personalization: **human in the** learning **loop**



Safety

Control based on a model of robot dynamics



Collaborative

From associations to understanding

Associative learning (perception ⟶ motion)

Attaching **semantics** to perceptions of objects and situations

Reasoning about **functionalities** and goals

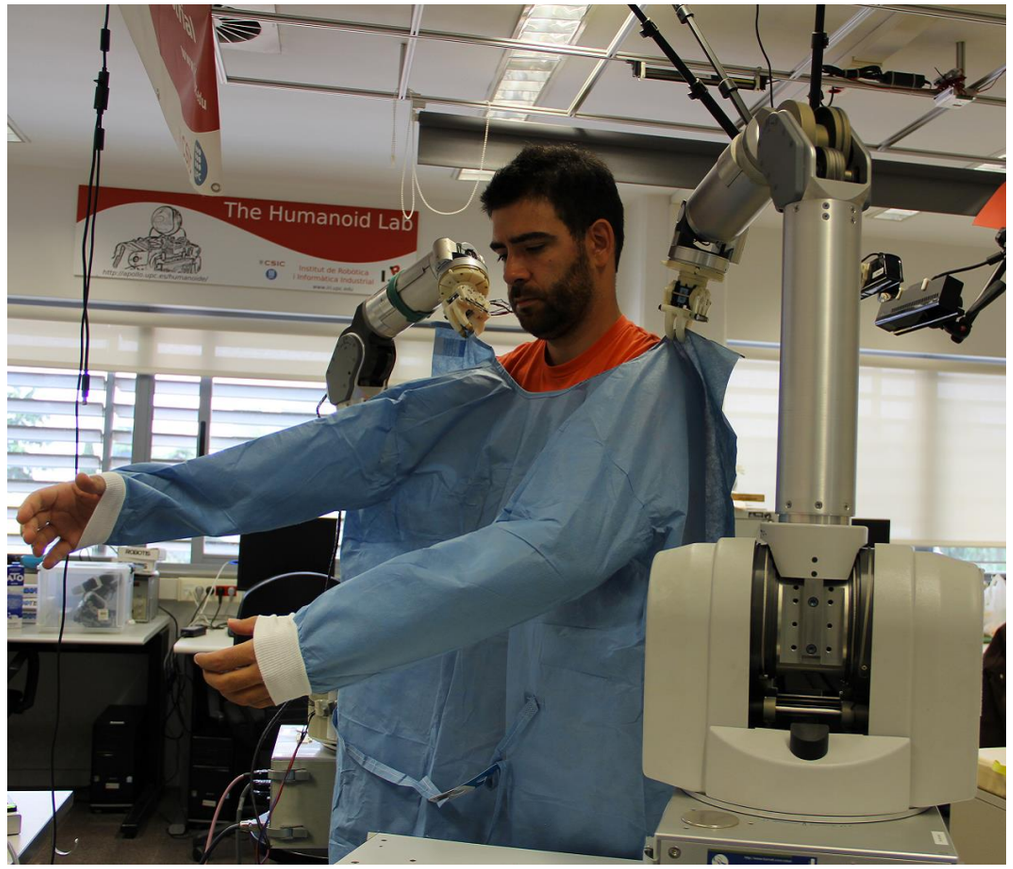


Table of contents

1a. Industrial Robotics ⟶ Assistive Robotics

1b. Clothing assistance: the CLOTHILDE project

2a. Ethic and social implications ⟶ Roboethics

2b. Education/debate through Science Fiction

Research challenges for AI

Robotic AI meets the Humanities

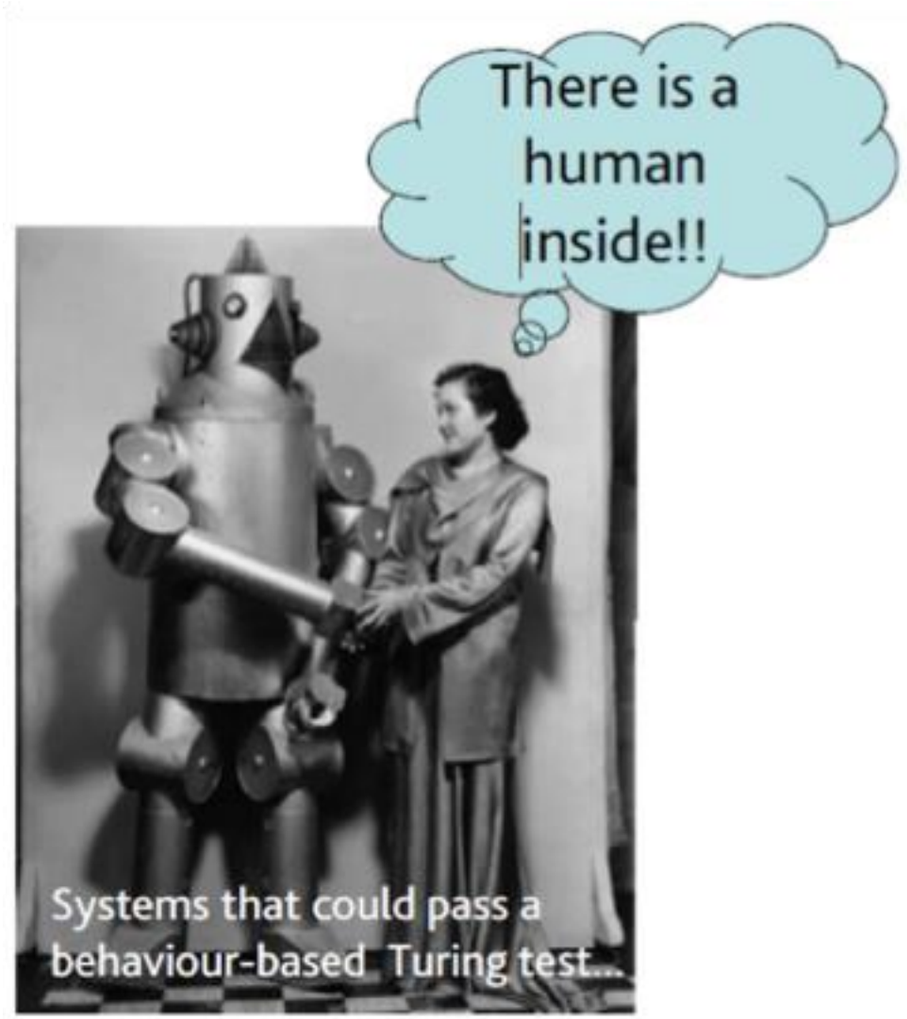
Ethical and social implications

Robotic assistants

Issues shared with other technologies:

- military usage
- incidence on the job market
- legal liability
- privacy
- digital gap
- …

Robot assistants raise new issues in entering the domain of **human decision-making, feelings, and relationships**.



Roboethics

Subfield of applied ethics studying both the **positive and negative implications of robotics for individuals and society**, with a view to inspire the moral design, development and use of so-called intelligent/autonomous robots, and help prevent their misuse against humankind.

[Veruggio *et al.* 2011]

Three aspects:

1. Human ethics applied to robotics

2. Codes of ethics embedded in the robots themselves (“machine ethics”)

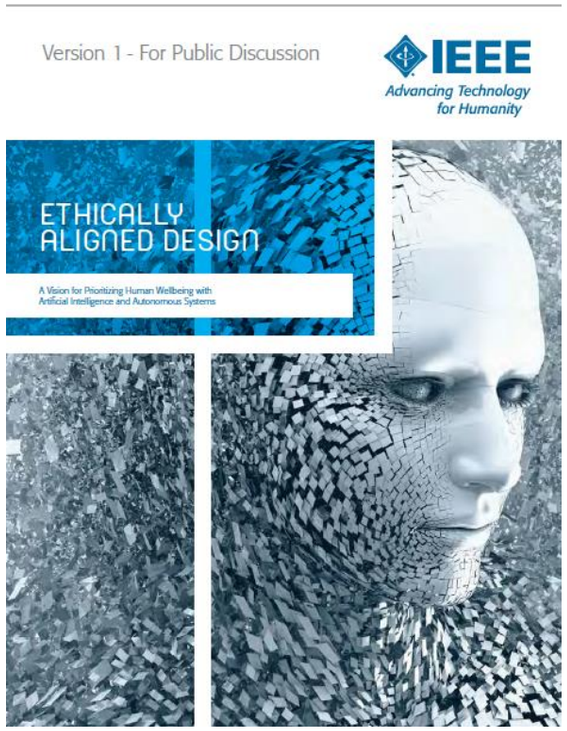
3. Ethics that would emerge from a potential future consciousness of robots

Roboethics

Regulations and standards

February 2017: The European Parliament's Plenary Session adopted the JURI report on **Civil Law Rules on Robotics (2015/2103(INL))**

- **RoboLaw project** "Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics": deliverable D6.2 - Guidelines on Regulating Robotics (2014)
- **euRobotics project** "The European Robotics Coordination Action": deliverable D3.2.1- Ethical Legal and Societal issues in Robotics (2012)



2016-18: IEEE Standards Association opened a draft on **Ethically Aligned Design** to public discussion.

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

An incubation space for new standards and solutions, certifications and codes of conduct, and consensus building for ethical implementation of intelligent technologies

Roboethics

Education and dissemination

A Cooperative Project of

Association for Computing Machinery
Advancing Computing as a Science & Profession

IEEE

IEEE computer society

The ACM/IEEE Computer Science Curricula consists of 18 knowledge areas, one of which is: "Social Issues and Professional Practice" that includes courses related to **Ethics in Technology**

Teach with Movies

LESSON PLANS BASED ON MOVIES & FILM!

Search TWM

SNIPPET LESSON PLAN FOR:

Robot Ethics Using Clips from Robot and Frank

Table of contents

1a. Industrial Robotics ⟶ Assistive Robotics

1b. Clothing assistance: the CLOTHILDE project

2a. Close to humans: Ethic and social implications

2b. Education/debate through Science Fiction

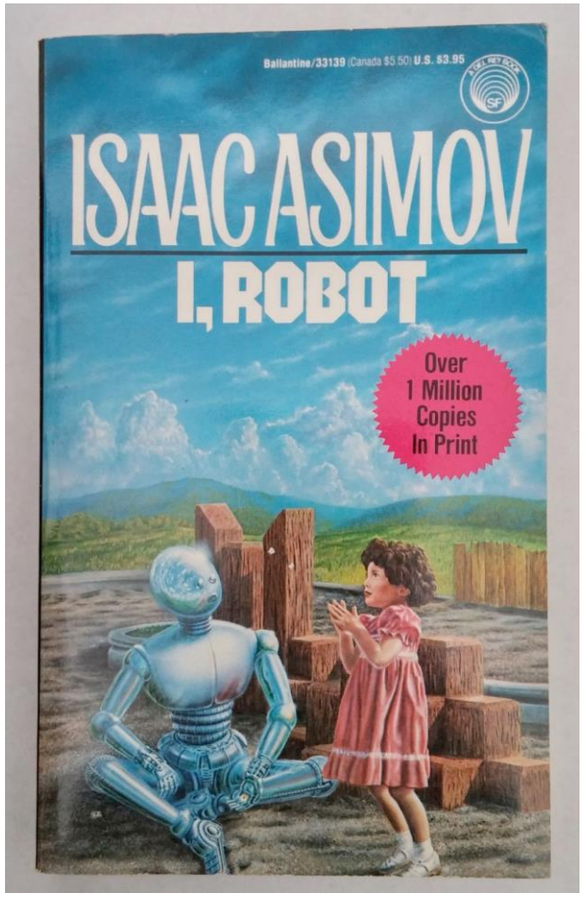
Research challenges for AI

Robotic AI meets the Humanities

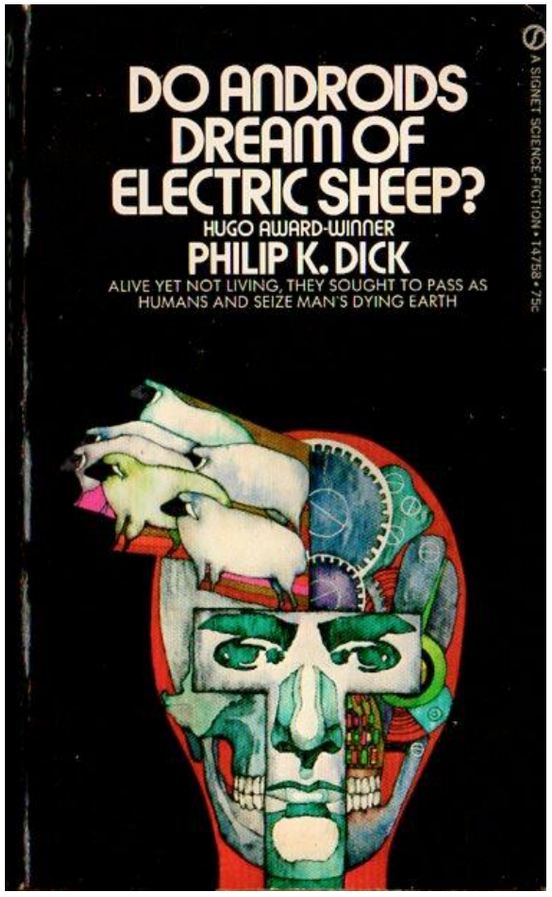
Roboethics

Education/dissemination based on Science Fiction

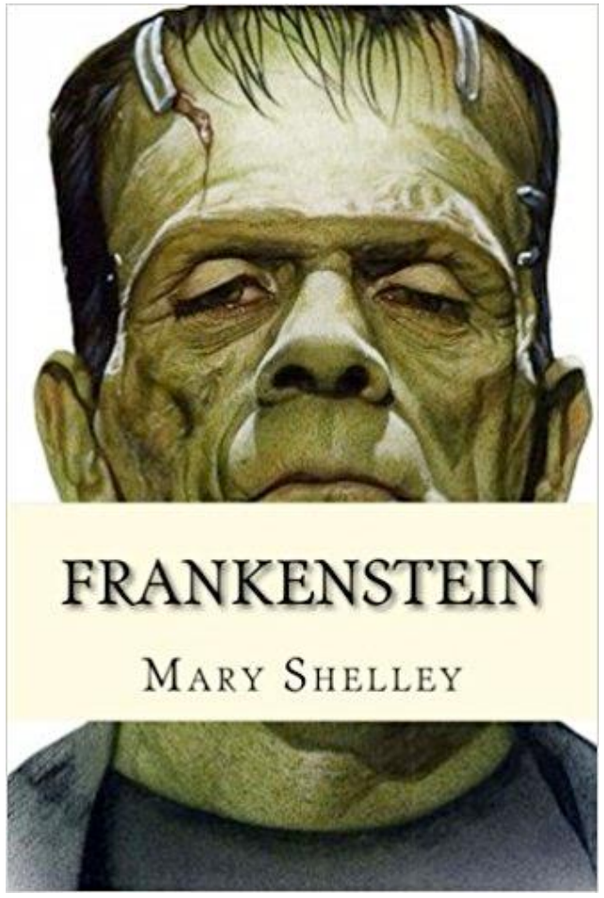
When trying to establish an ethical debate, disseminate concepts to the general public, or teach a course on roboethics, classic science fiction stories are often used to exemplify possible future conflicting situations.

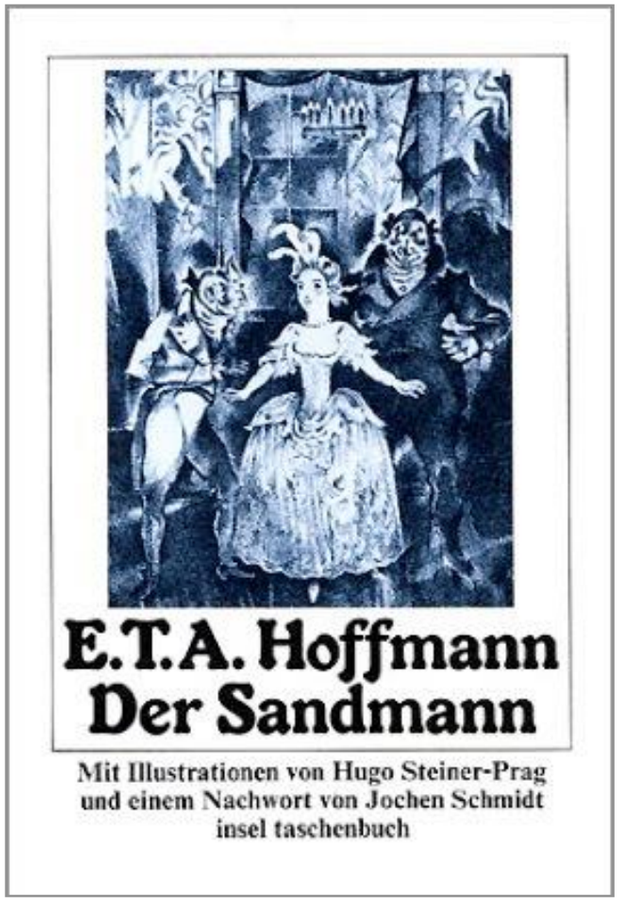


Robot nannies



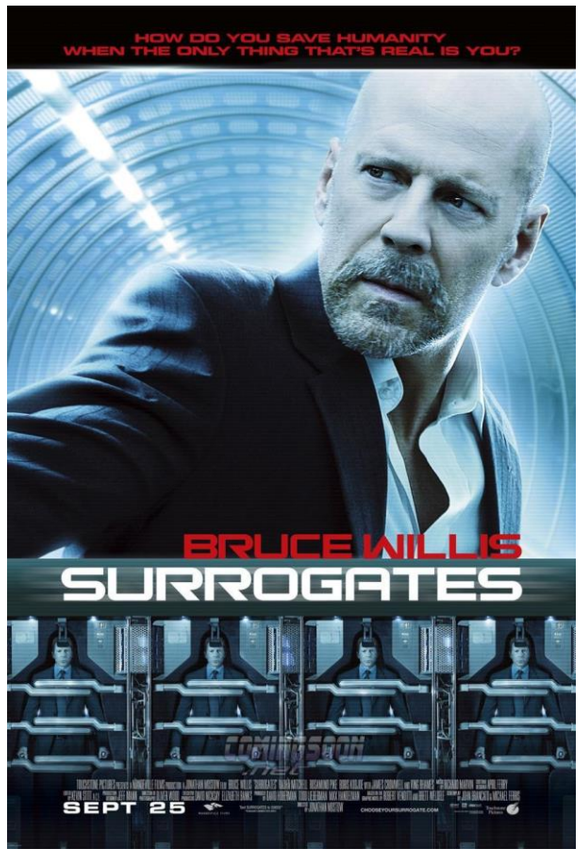
Humanoid replicas

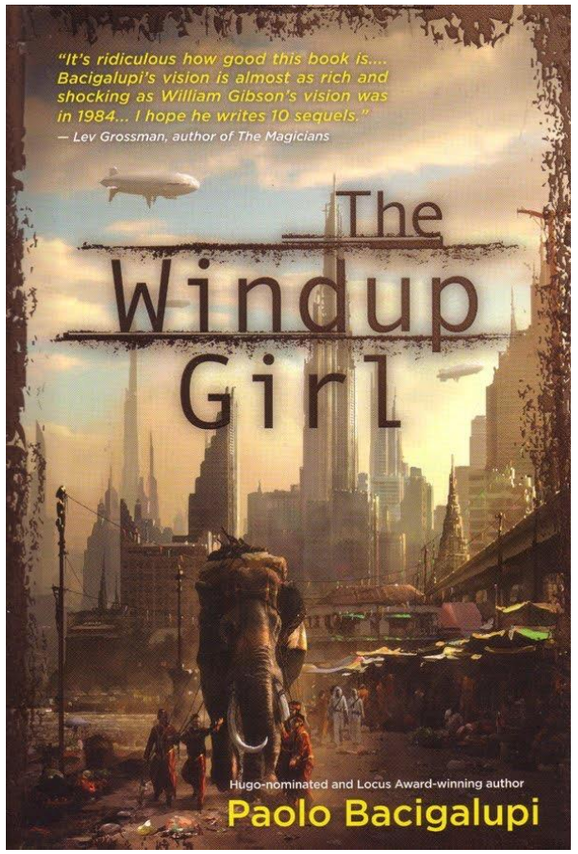




Emotional surrogates

Modern Science Fiction related to Roboethics



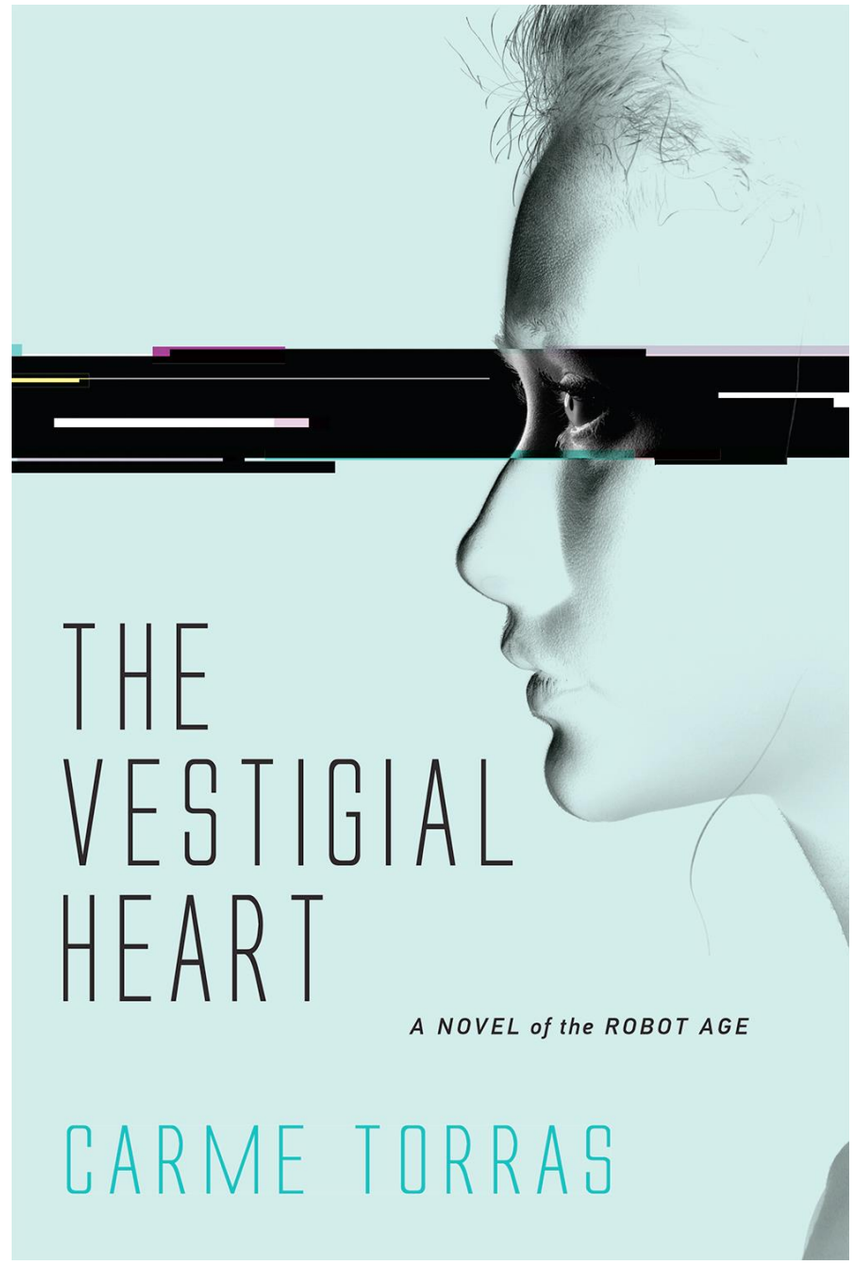


"It is the relationships that we have constructed which in turn shape us"

Robert C. Solomon
"The Passions"

Modern Science Fiction related to Roboethics

Modern Science Fiction related to Roboethics



https://mitpress.mit.edu/books/vestigial-heart

Modern Science Fiction related to Roboethics

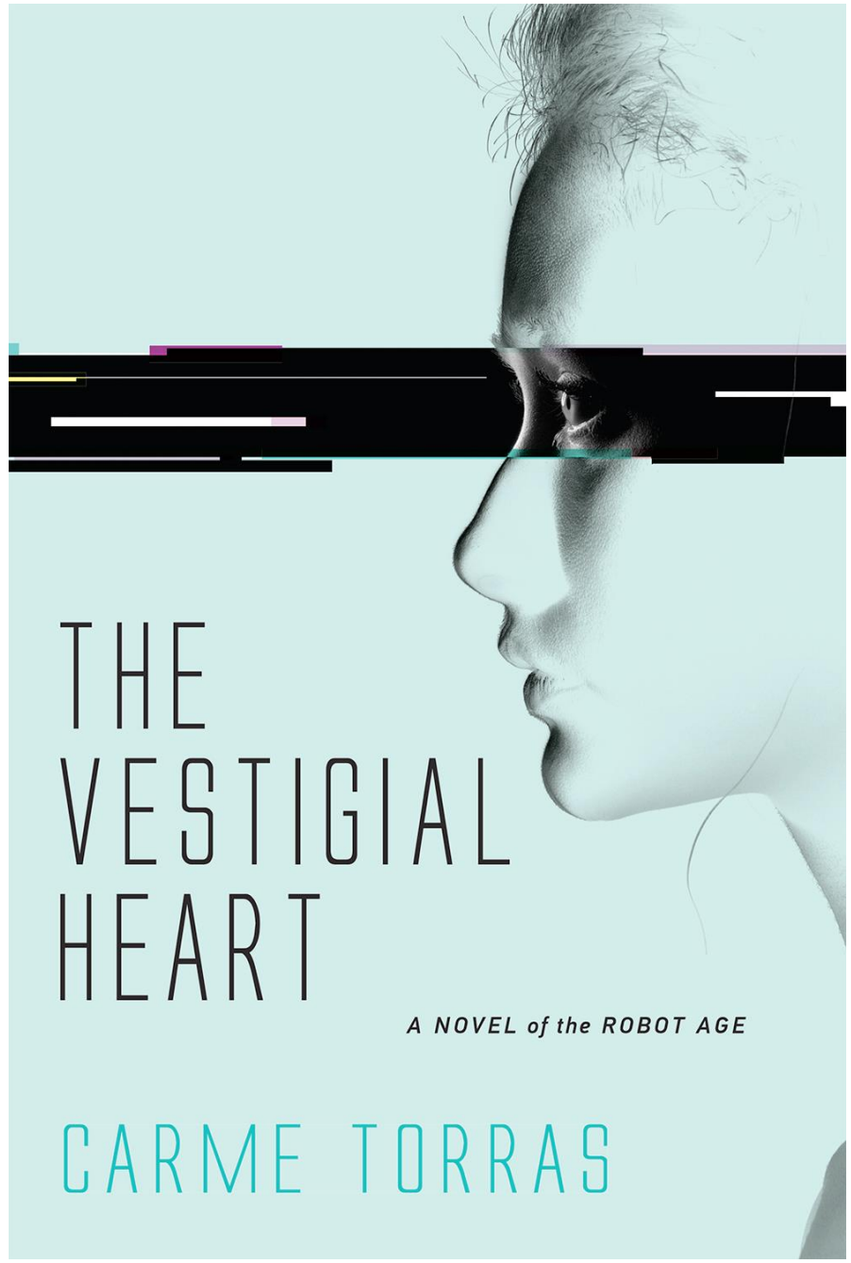
Four items:

- A **novel** about a future society in which people rely on personal-assistant robots to navigate daily life.

- An **appendix** with 24 ethics questions raised by the novel, as well as hints to trigger a debate.

- An **online teacher's guide** for 6-8 sessions on "Ethics in Social Robotics and AI" following the chapters in the novel and including scholarly references for further reading.

- A **100-slide presentation** that teachers can use and extend as desired.

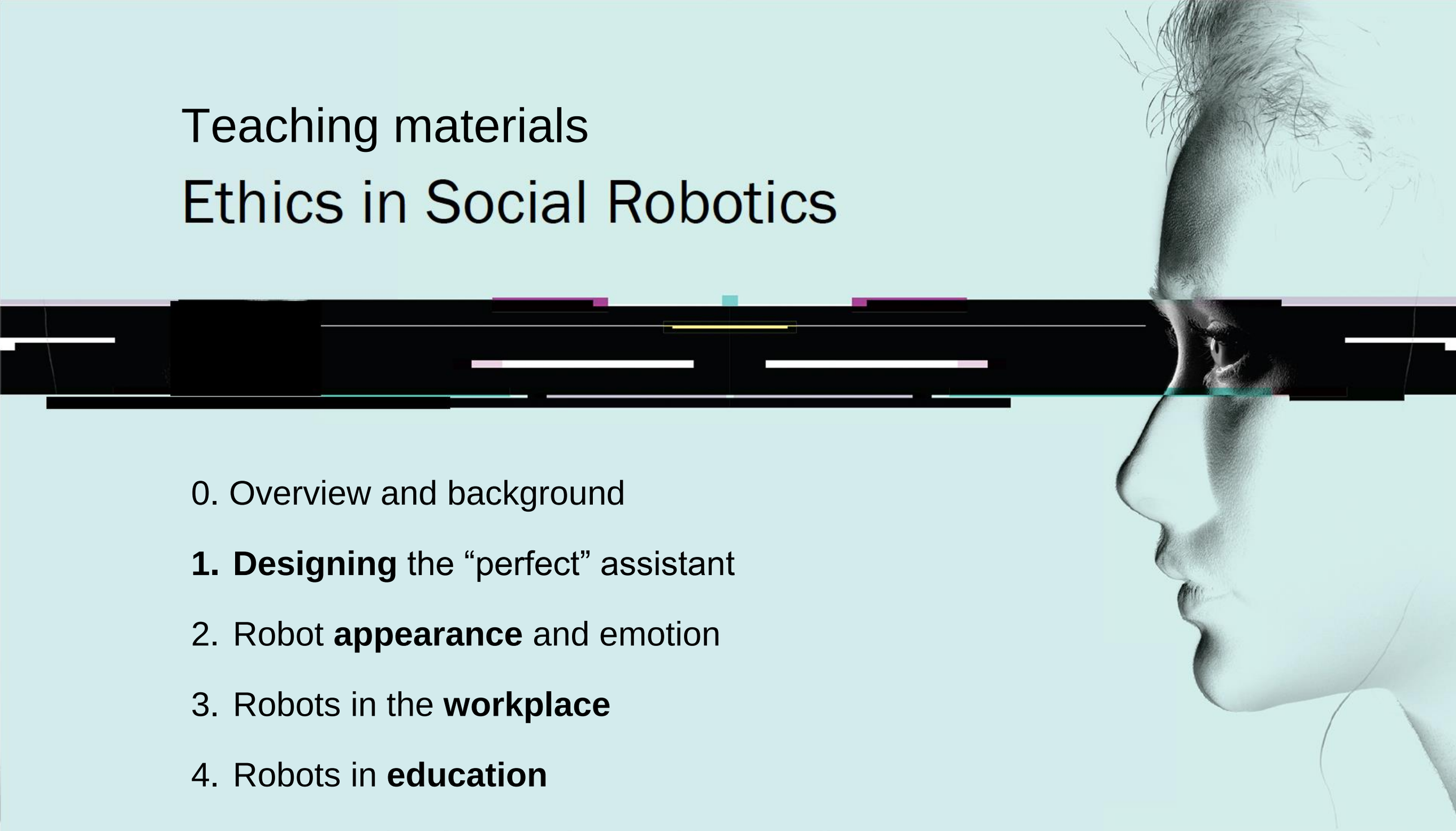


The MIT Press

https://mitpress.mit.edu/books/vestigial-heart

Teaching materials

Ethics in Social Robotics



0. Overview and background
1. **Designing** the “perfect” assistant
2. Robot **appearance** and emotion
3. Robots in the **workplace**
4. Robots in **education**
5. Human-robot **interaction** and human dignity
6. Social **responsibility** and robot morality
7. Bibliography and initiatives to follow up

The MIT Press

0.2. Ethical theories relevant to robotics

- **Utilitarianism or Consequentialism** — maximizing the number of people that enjoy the highest beneficial outcomes
- **Deontologism** — acting only according to maxims that could become universal laws
- **Virtue ethics** — relying on the moral character of virtuous individuals
- **Social justice** — all human beings deserve to be treated equally and there must be a firm justification in case of mistreatment
- **Common goods** — living in a community places constraints on the individual
- **Religious ethics** — norms come from a spiritual authority
- **Information ethics** — policies and codes for governing the creation, organization, dissemination, and use of information

[Sullins 2015]

0.2. Ethical theories relevant to robotics: Hybrid ethics

Since no single theory is appropriate for addressing all ethical issues arising in the design and use of robots, we take a pragmatic option:

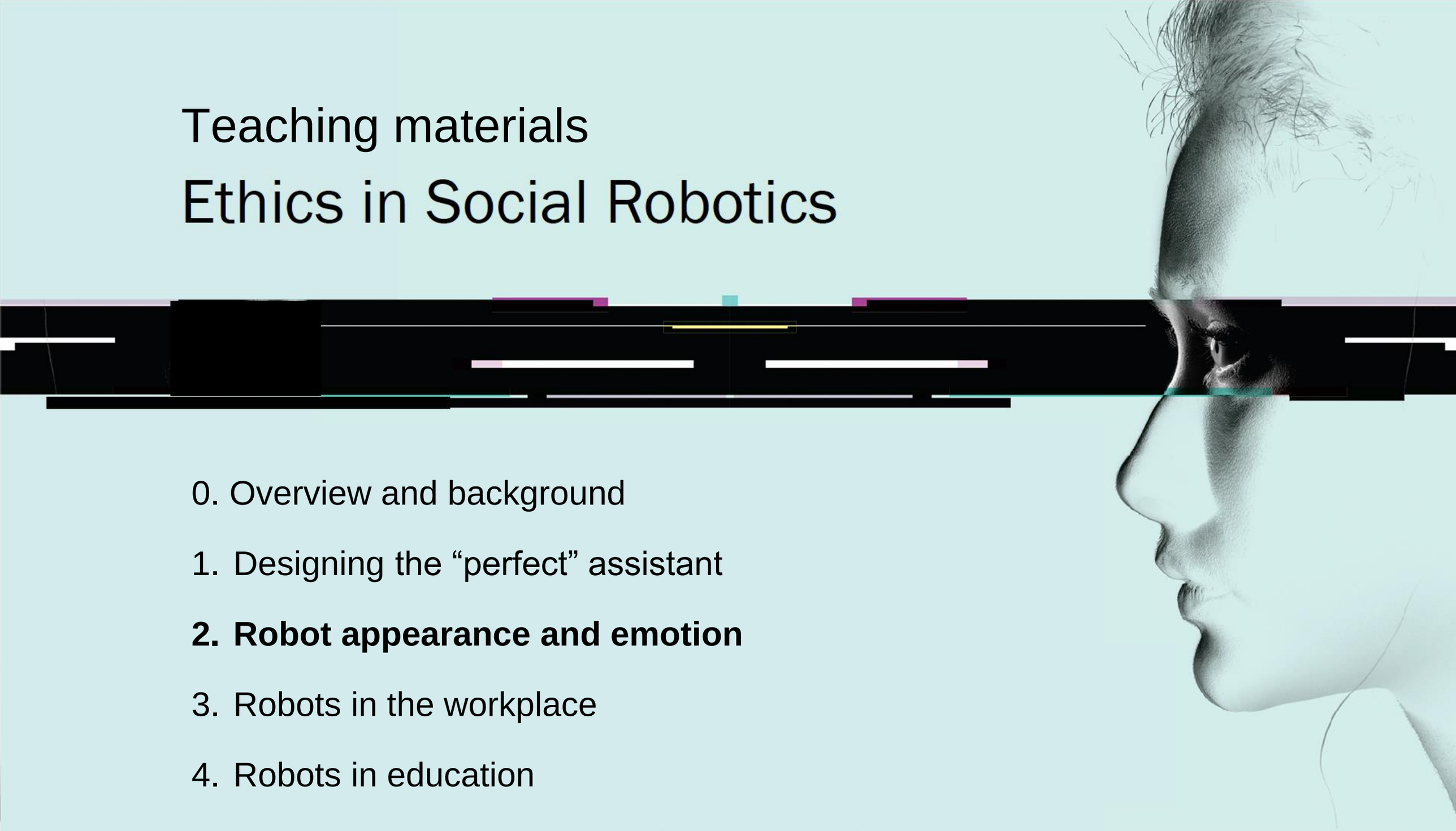
Hybrid ethics combines:

- Top-down theories — those applying rational principles to derive norms, and
- Bottom-up theories — those inferring general guidelines from specific situations.

[Wallach and Allen 2018]

Teaching materials

Ethics in Social Robotics

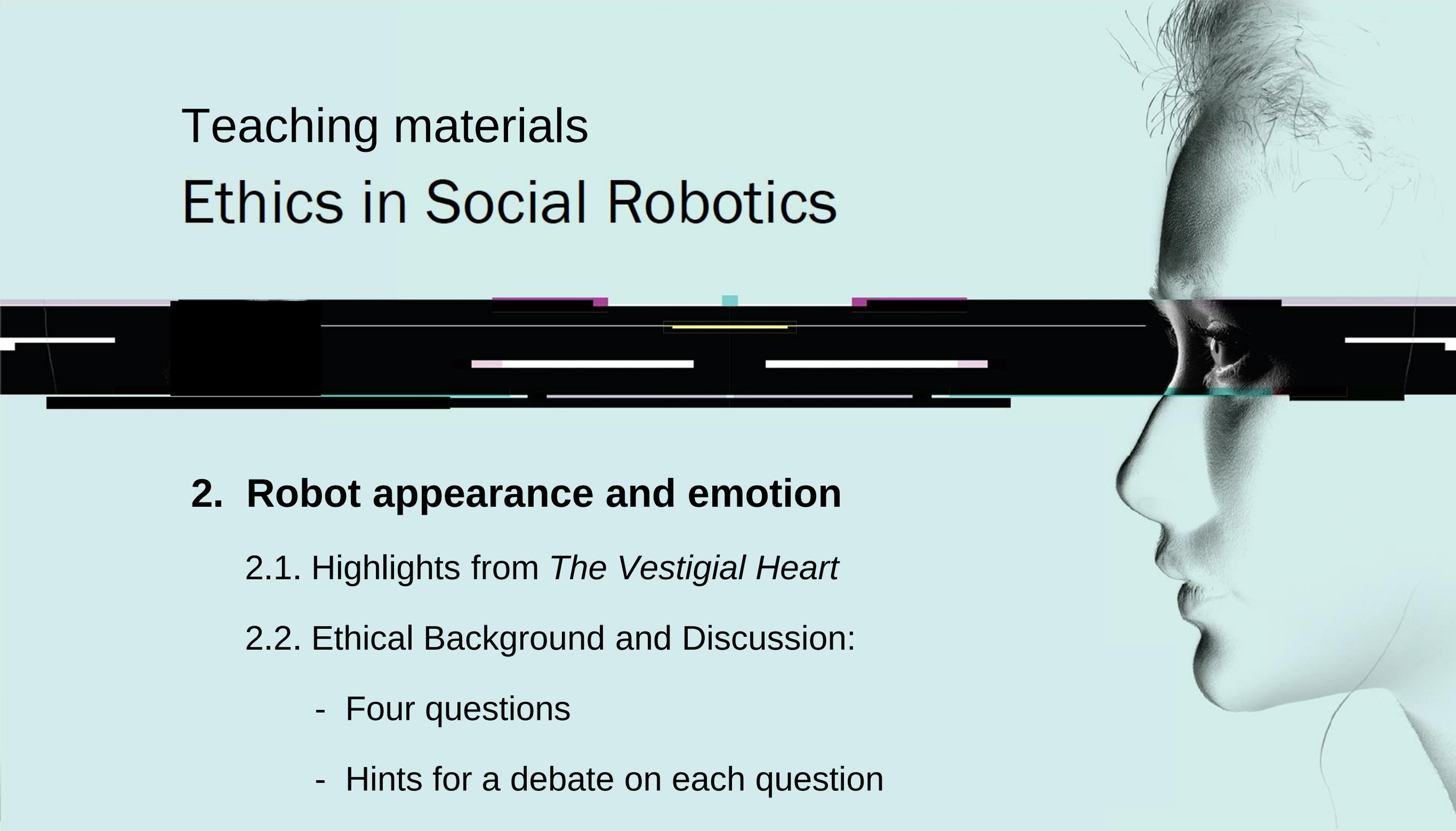


0. Overview and background
1. Designing the “perfect” assistant
2. **Robot appearance and emotion**
3. Robots in the workplace
4. Robots in education
5. Human-robot interaction and human dignity
6. Social responsibility and robot morality
7. Bibliography and initiatives to follow up

The MIT Press

Teaching materials

Ethics in Social Robotics



2. Robot appearance and emotion

2.1. Highlights from *The Vestigial Heart*

2.2. Ethical Background and Discussion:

- Four questions
- Hints for a debate on each question

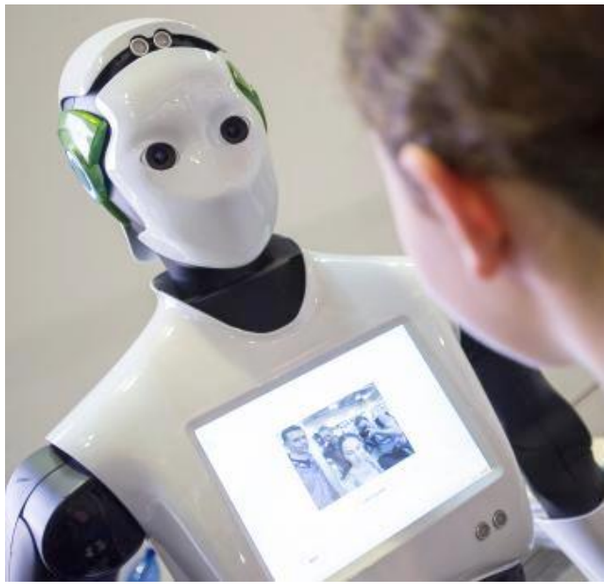
2.3. Revisiting Issues

2.4. Scholarly References for Further Reading

The MIT Press

2.1. Highlights from *The Vestigial Heart*

Chapters 9/12 - Celia



As a birthday present, Lu [adoptive mother] gave me a robot. [..] it has a kind of head with no nose, mouth or ears, it just has two cameras, and a screen embedded in its chest. It's called ROBbie.

[..] Celia, touched by the words, looks for his eyes: no friend had ever sworn their loyalty so convincingly, but two black holes bring her back down to earth. Though not entirely.

[..] she watches the robot out of the corner of her eye and it pleases her to see his dignified posture, gently swinging his strong, shiny arms. It feels good to walk along beside him, she feels protected, she can trust him. What does it matter that he doesn't have eyes, people don't look at each other anymore anyway."

At the Disasters stand, Leo is puzzled by a realistic mechanical baby. [..] What woman could resist the charm of a baby that smiles when she coos at it, that she can cuddle at will while watching her favorite program, that recognizes her voice and crawls along behind her, flattering her with sweet noises? Well no sir, the product didn't take off, almost certainly because it's too much like the real thing, déjà vu.

Robot appearance and emotion

Debate stemming from Chapters 9, 10 and 12 in the novel

Appearance strongly influences people's attitude towards robots: the more anthropomorphic the robot, the more positive and empathetic the human response.

[Riek *et al.* 2009]

However, the relation doesn't grow unlimited; on the contrary, a point is reached where excessive similarity of the robot to a human causes distress and provokes a sudden repulsion; this is known as the "uncanny valley" effect.

Celia feels attached to her robot ***ROBbie*** because of its loyal, trustworthy and predictable behavior, which is enforced by its undeceiving machine appearance.

Leo realizes that a too-close similarity to a human being can doom a robot product.

Robot appearance and emotion

2.2. Questions

2.A - How does robot appearance influence public acceptance?

2.B - What are the advantages and dangers of robots simulating emotions?

2.C - Have you heard of/experienced the “uncanny valley” effect?

2.D - Should emotional attachment to robots be encouraged?

Robot appearance and emotion

2.A - How does robot appearance influence public acceptance?

Anthropomorphism provides very powerful physical and social features that will be implemented in assistive robots as they ease communication with users. However, «Are we constraining a robot to become too animalistic (including humanistic) that we miss how it can constructively contribute to our way of life?» [Duffy 2003]

An ethics of appearance and human good entails listening to people's experience and focusing on how human-robot interactions may enrich our emotional life in a possibly different and complementary way to human-human relationships.

[Coeckelbergh 2009]

Exploring alternative robotic morphologies:

- robotize some everyday objects to respond to human intentions and emotions [Sirkin and Ju 2014]
- innovative prototyping methods for designing socially situated embodiment [Sabanovic *et al*. 2014]

Generic consideration: avoid sexist, ableist, racist and ethnic robot morphologies.

Robot appearance and emotion

2.B - What are the advantages and dangers of robots simulating emotions?

Emotion expression by a robot plays a prominent role in social, face-to-face interactions with people. [Breazeal 2003]

Advantage: In a search and rescue setting, nonverbal expressions of negative mood and fear by the robot improve the participants' compliance with its request to evacuate, causing them to respond earlier and faster. [Moshkina 2012]

Deception is an important danger of affective simulation as it infringes autonomy, because misinforming a person about the alternatives that are open prevents him or her from choosing rationally between them. In this regard, vulnerable users (e.g., children and elderly people) deserve especial attention. [Cowie 2014]

Note the difference with a robot capturing the emotional state of the user and acting accordingly, which can be very handy in some healthcare situations.

Although the robot is only expressing a simulated emotion, the feelings it evokes in people are real and may be strong. [Turkle 2007]

Robot appearance and emotion

2.C - Have you heard of/experienced the “uncanny valley” effect?

This name follows from the shape of the curve representing human attitude as a function of robot anthropomorphism, which grows steadily up to a point where it falls down into a profound valley. Just picture the reaction triggered by the fictional robot WALL-E and compare it to that produced by the initial geminoids at Osaka University, which look almost human but not quite, causing a creepy impression.

The uncanny/creepy impression can help understand what human qualities robots are missing and thus trigger research to improve their communication abilities.

[MacDorman and Ishiguro 2006]

What seems to trigger repulsion is the lack of coherence between the different elements that mediate human-robot communication: appearance, expression, language, speech, gestures, posture, motion, responsiveness to attentive and motivational cues, interaction speed, turn taking in dialogs, synchronization, etc.

«Authenticity is for us what sex was to the Victorians: taboo and fascination, threat and preoccupation.»

[Turkle 2007]

Robot appearance and emotion

2.D - Should emotional attachment to robots be encouraged?

Aliveness seems to be the key factor for the development of affective ties with a mechanical creature. Participants in an experiment hesitated three times as long to switch off an agreeable and intelligent robot (perceived as a living creature) as compared to a non-agreeable and unintelligent one. [Bartneck *et al.* 2007, 2009]

Experience episodes categorized into three dimensions: visceral, behavioral, and reflective. [Norman 2004] Vulnerable groups tend to base attachment on the former dimensions, differing from adults in full judgment. [Weiss *et al.* 2009]

Opposite views:

It is completely normal that people fall in love with artificial companions. [Levy 2010]

Machines should always be just servants that you can switch off whenever you like. [Bryson 2010]

Robot appearance and emotion

2.D - Should emotional attachment to robots be encouraged? (cont.)

Examples of potential benefits of attachment:

- making robot nudging more compelling
- providing company/sex to adults who deliberately make this choice.

The main danger is the user's social isolation, which can derive from:

- family and friends eluding their responsibilities once the robot takes care
- the seductions of the robot itself ('lotus eater' problem) [Cowie 2014]
  Living and engaging with robots will be so easy that human relationships will be discouraged because they would just seem too hard. [Turkle 2007]

Note of caution: beyond the discussion on encouraging or discouraging the formation of emotional bonds, roboticists should be aware that some bonding will be inevitable regardless of the morphology of the robot. [Riek *et al.* 2009]

Robot appearance and emotion

2.3. Revisiting issues

"I [Celia] have been more shocked by the kids, and even some things Lu does, than by ROBbie. For the robot, everything follows a series of rules, it'll never surprise me with anything inappropriate."

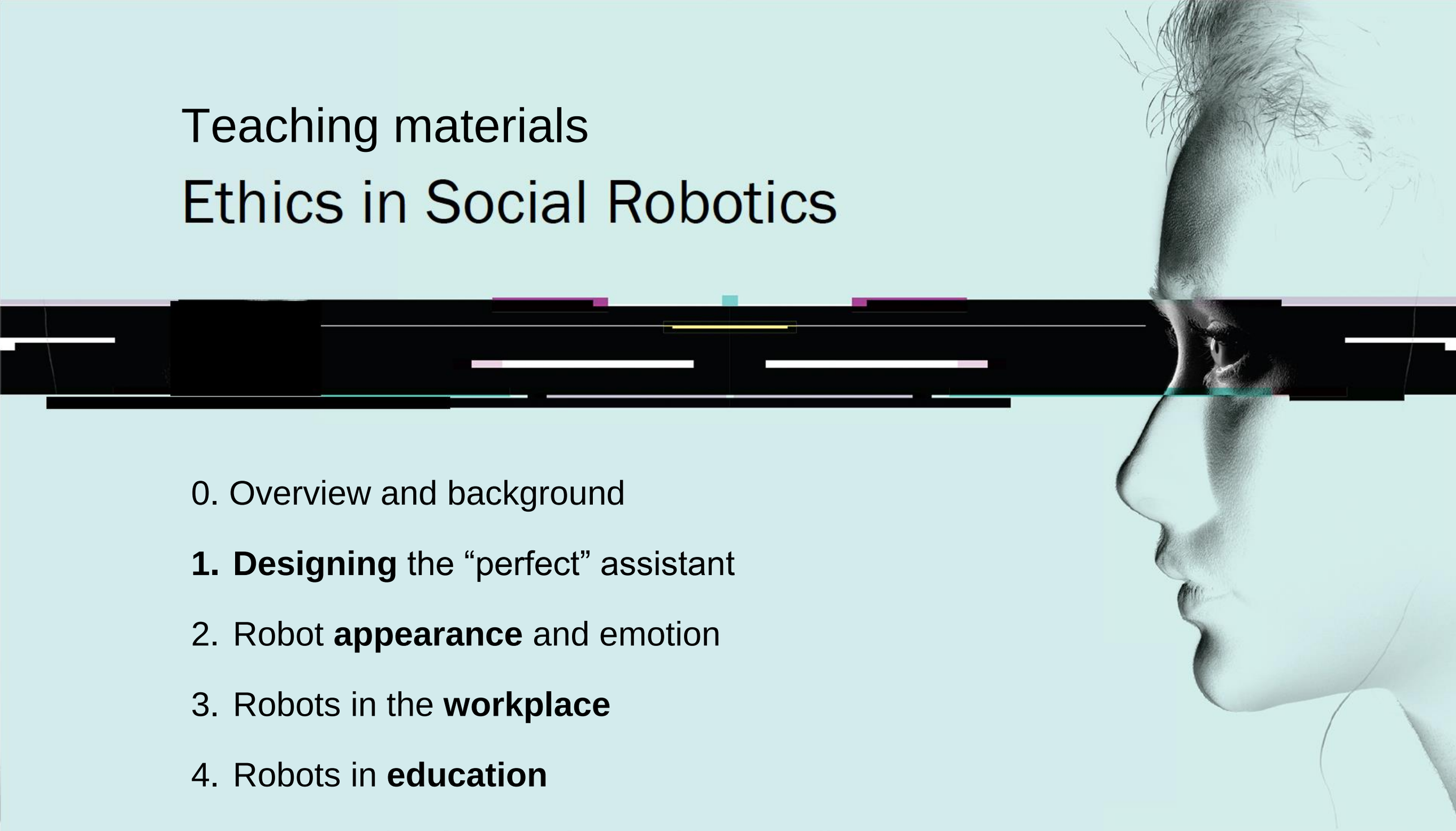
Transparency (1.C): ***Celia*** likes that ***ROBbie*** has a more predictable behavior than her classmates and her adoptive mother, since it has to follow rules and can't shock her with nonsense.

"It feels good to walk along beside him [ROBbie], she feels protected, she can trust him."

Trust (1.A): ***Celia*** feels protected by her robot, which she sees as a faithful companion that she can trust.

Teaching materials

Ethics in Social Robotics



0. Overview and background
1. **Designing** the “perfect” assistant
2. Robot **appearance** and emotion
3. Robots in the **workplace**
4. Robots in **education**
5. Human-robot **interaction** and human dignity
6. Social **responsibility** and robot morality
7. Bibliography and initiatives to follow up

The MIT Press

Amazing future perspectives

What role will the human and the robot play in this “pas à deux” in which we are irremissibly engaged?

Carme Torras

torras@iri.upc.edu