# Working with [robots/humans] to make better [humans/robots]

Katie Winkle

# How will Humans and AI Interact in 5 Years?

What does good look like?

What can these interactions really do for us? What are the risks?

# How will Humans and 🤖 Interact in 5 Years?

*Robots as tangible AI*

What do we want vs what will we get?

Will people accept our robots? Use them or abuse them? (Over) trust them?

# Three Types of Human Robot Interaction

(1) Working with robots to make humans better

# Three Types of Human Robot Interaction

(1) Working with robots to make humans better

(2) Working with humans to make robots better

# Three Types of Human Robot Interaction

(1) Working with robots to make humans better

(2) Working with humans to make robots better

(3) Working with robots to make better humans

# Three Types of Human Robot Interaction

(1) Working with robots to make humans better

(2) Working with humans to make robots better

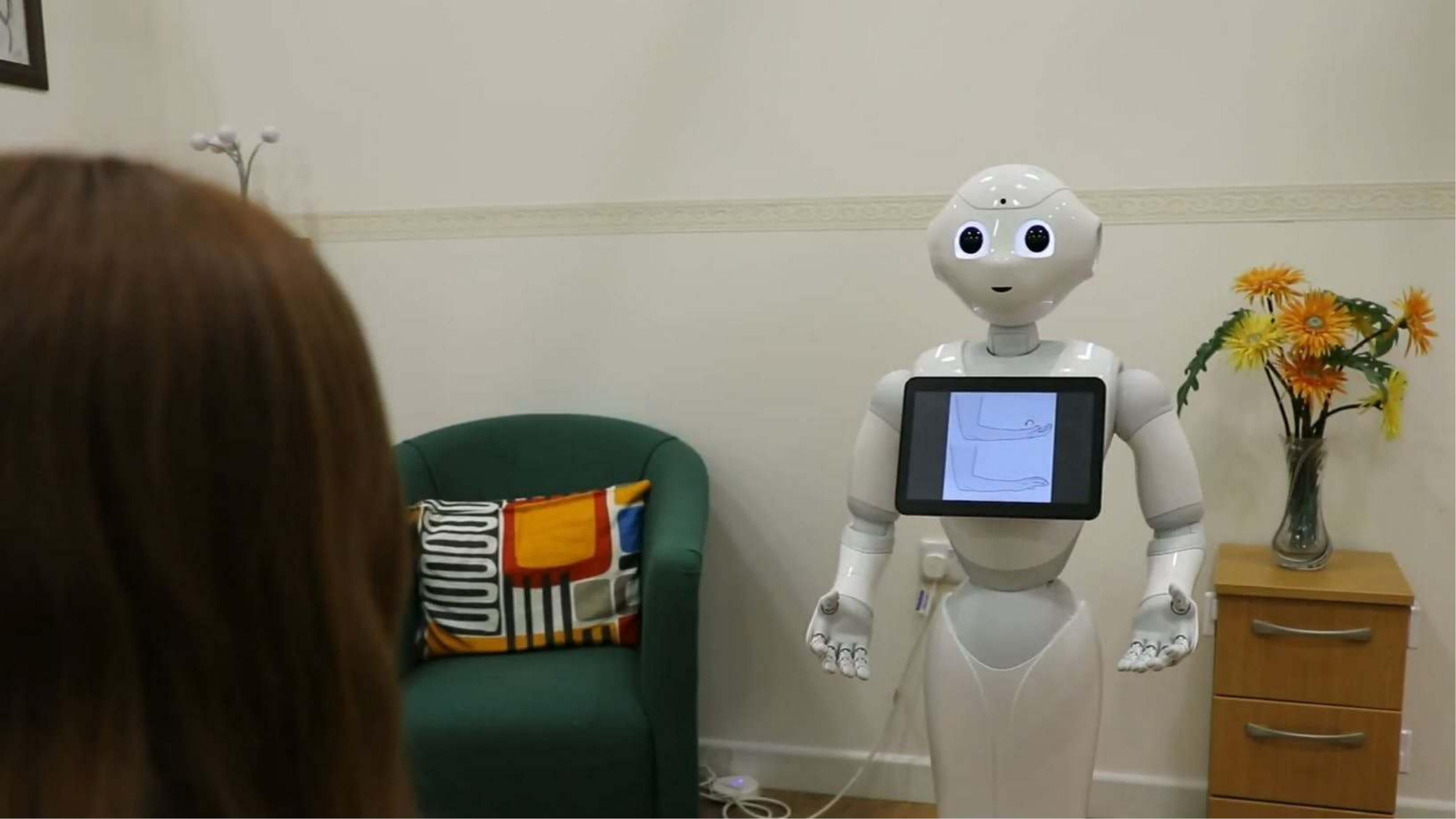
(3) Working with robots to make better humans

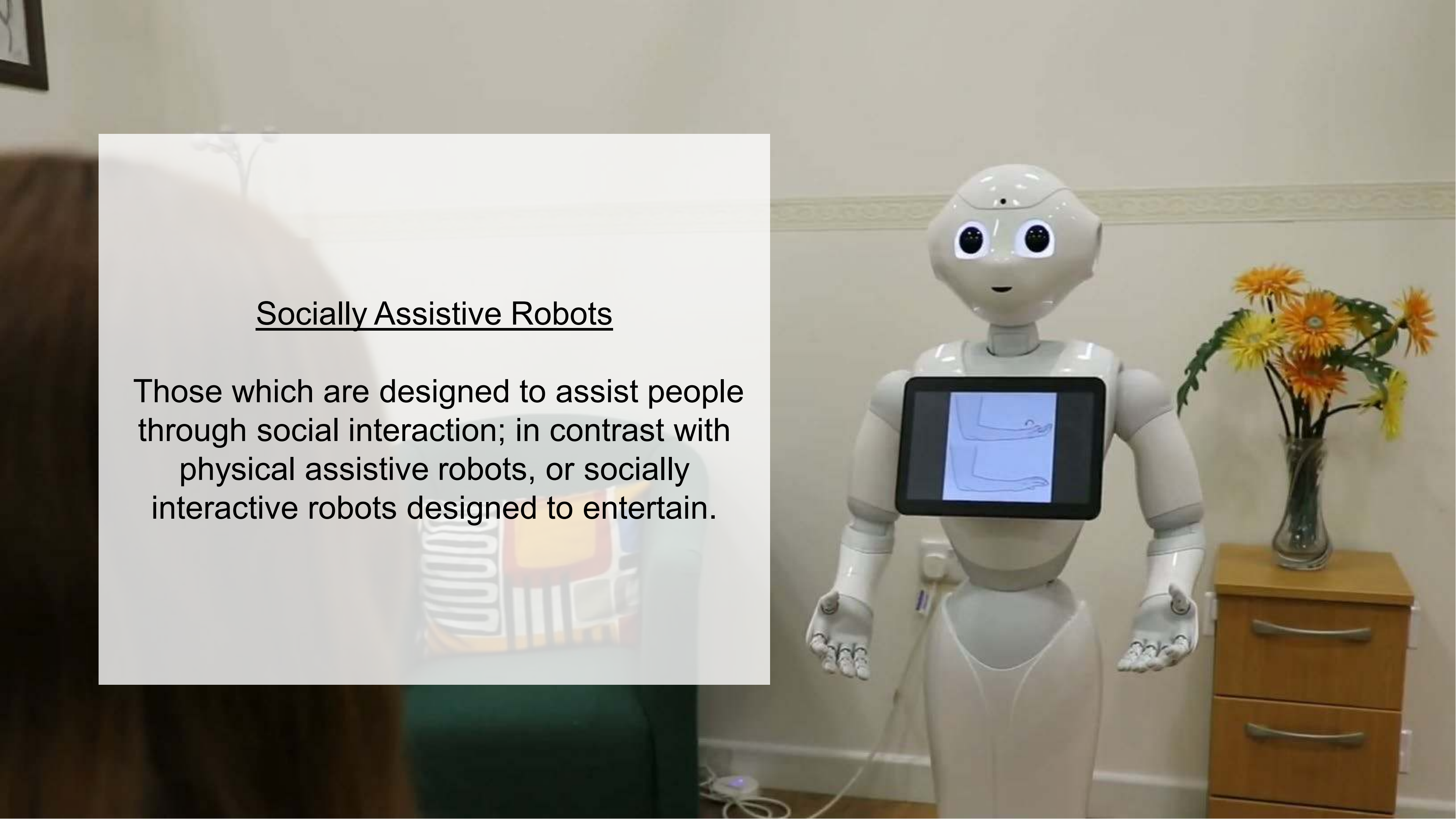
# Responsible Robotics = Effective Robotics

Key to my research philosophy is that 'responsible' approaches:

- complimenting not replacing human-human interaction
- ensuring diversity in/democratising robot development
- avoiding propagation of bias
- being mindful of the broader implications of tech. deployment

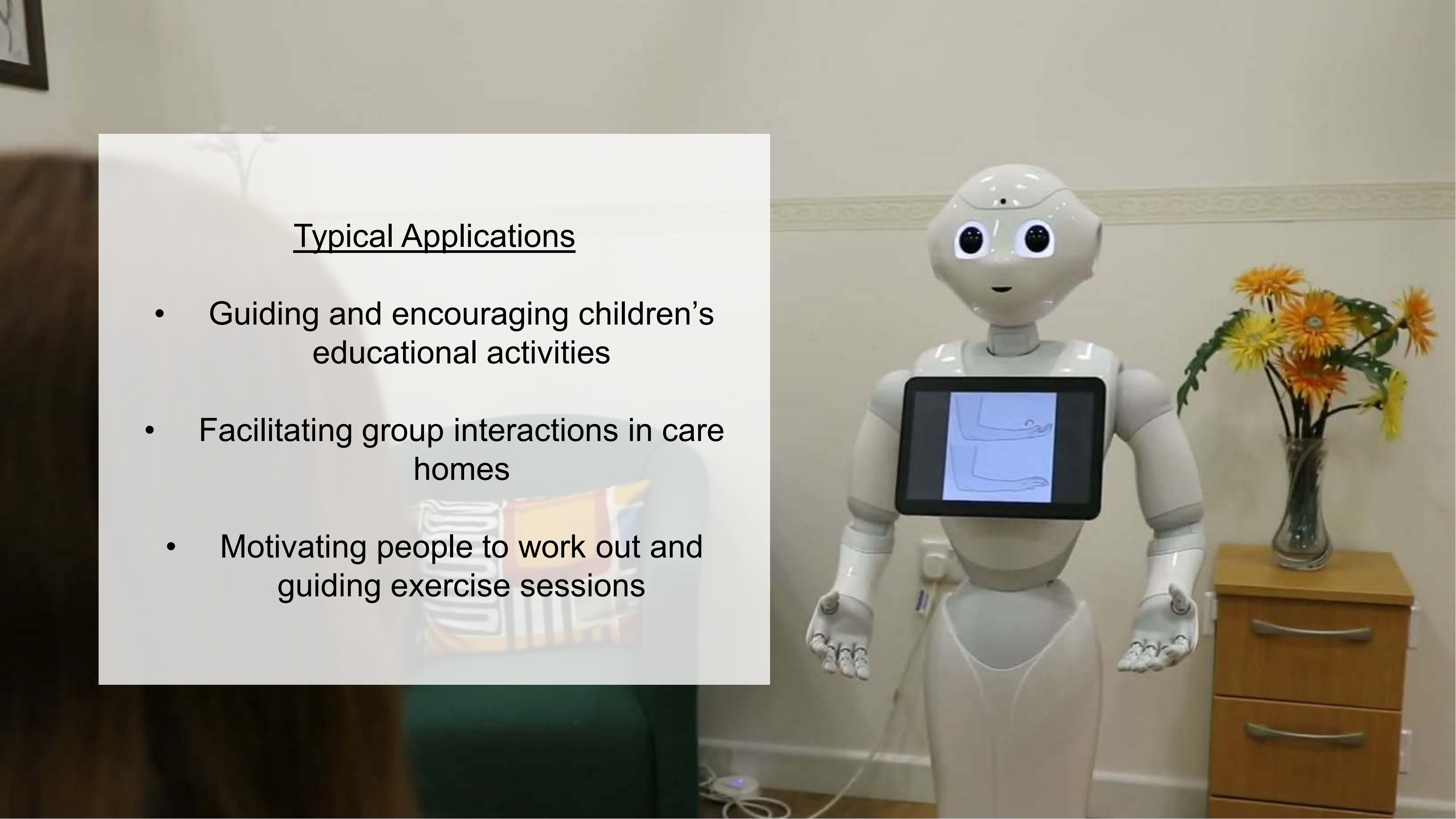
are fundamentally *good* approaches for building state-of-the-art *technical* systems.





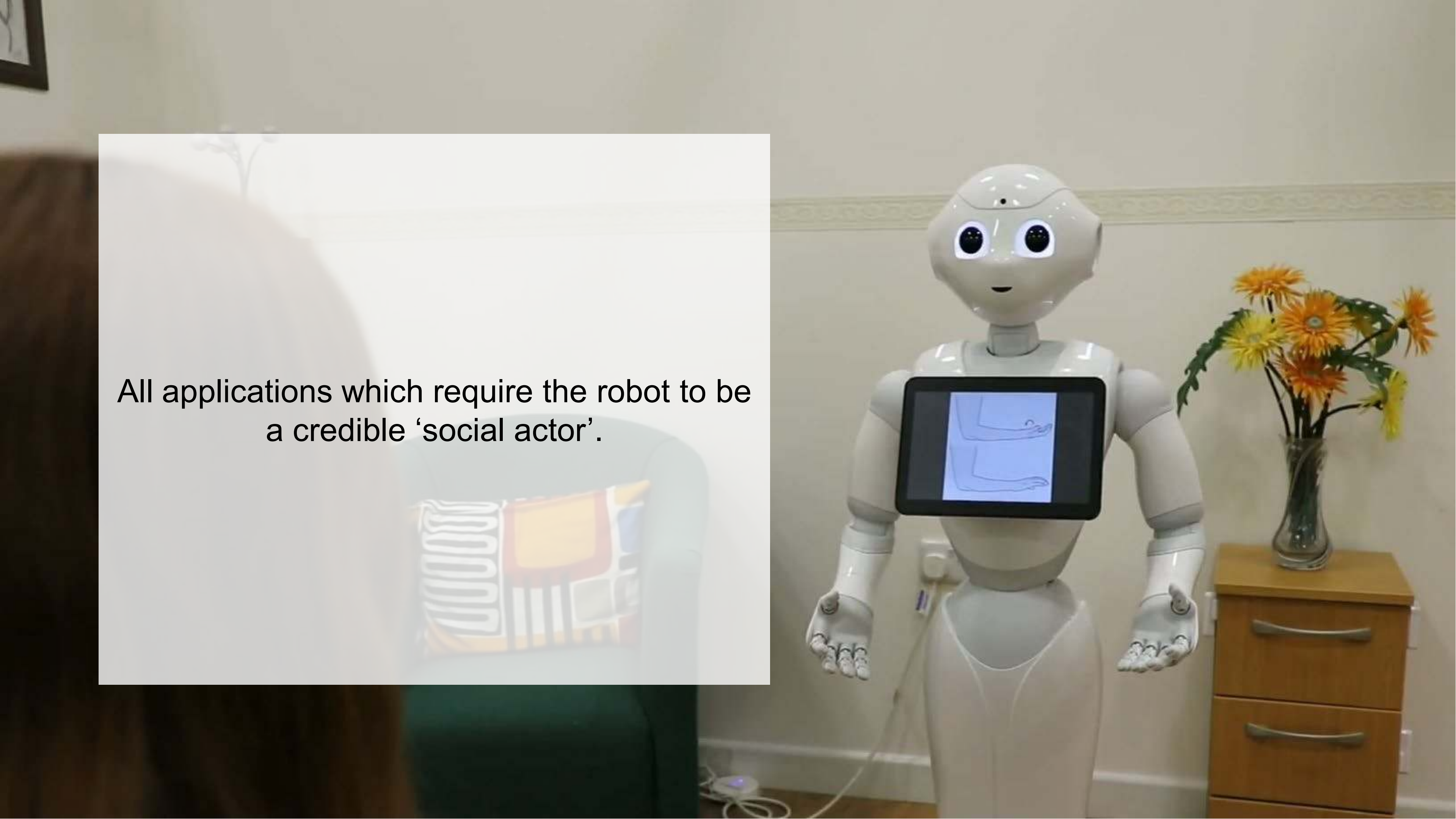
Socially Assistive Robots

Those which are designed to assist people through social interaction; in contrast with physical assistive robots, or socially interactive robots designed to entertain.



## Typical Applications

- Guiding and encouraging children’s educational activities
- Facilitating group interactions in care homes
- Motivating people to work out and guiding exercise sessions



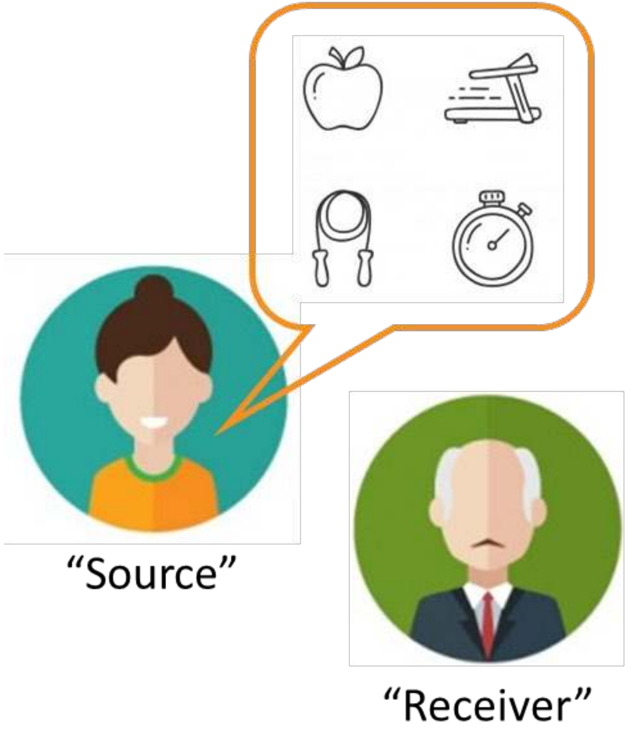
All applications which require the robot to be a credible ‘social actor’.

This part of training will only be effective…

…if there is also this rapport and social interaction underpinning it.

# The Elaboration Likelihood Model of Persuasion

A model of (human – human) persuasion that nicely explains this importance of ‘off-topic’ social interaction and rapport in motivation.



# The Elaboration Likelihood Model of Persuasion

Credible?
Likeable?

High Elaboration

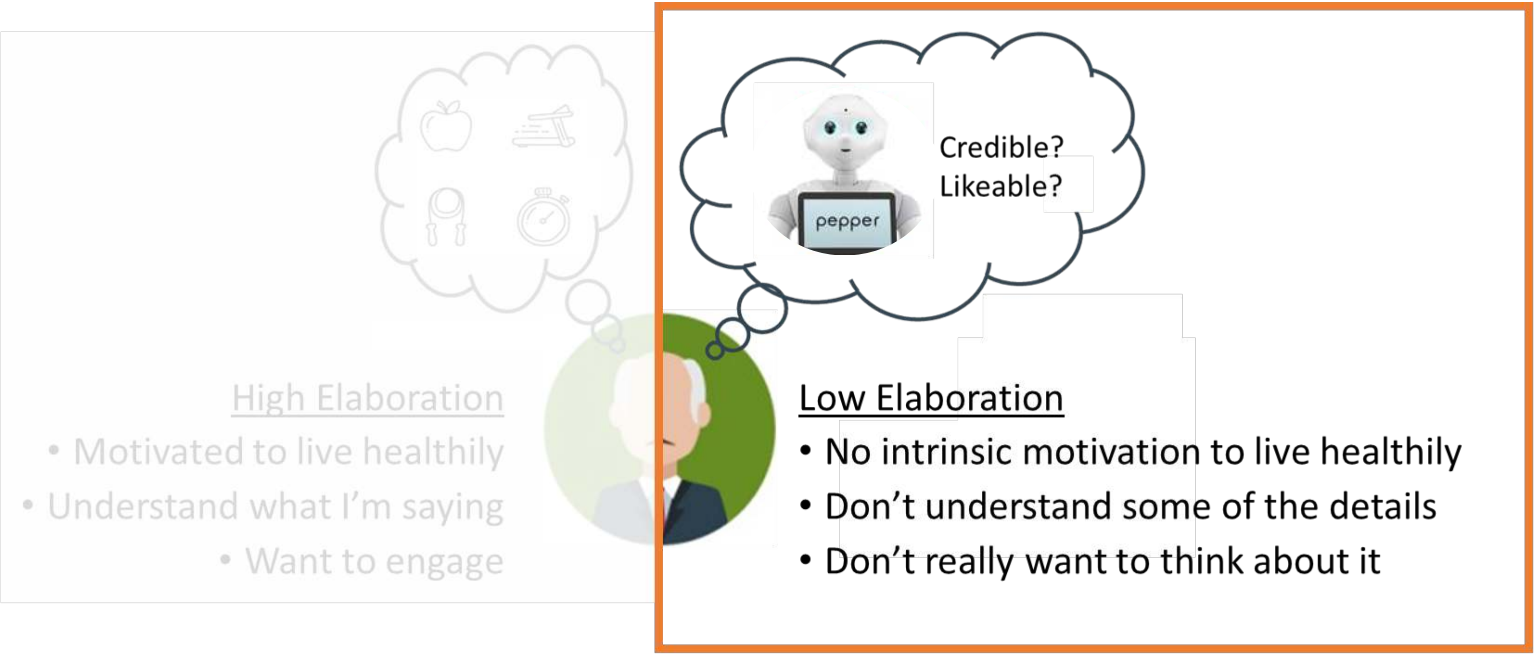
- Motivated to live healthily
- Understand what I'm saying
- Want to engage

Low Elaboration

- No intrinsic motivation to live healthily
- Don't understand some of the details
- Don't really want to think about it

Social assistance scenarios

# The Elaboration Likelihood Model of Persuasion



High Elaboration

- Motivated to live healthily
- Understand what I'm saying
- Want to engage

Low Elaboration

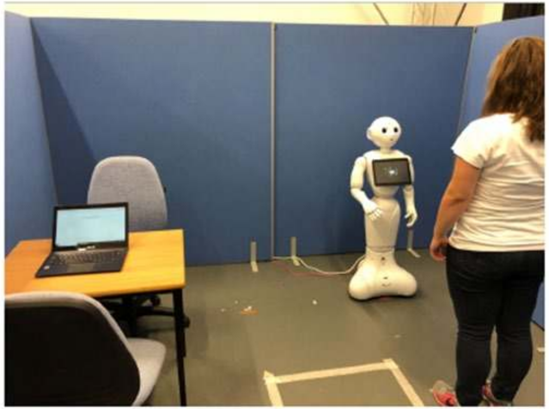
- No intrinsic motivation to live healthily
- Don't understand some of the details
- Don't really want to think about it

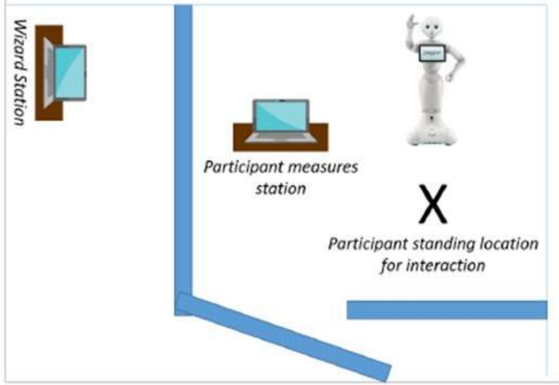
Social assistance scenarios

# An Assistive (Persuasive) Social Robot

Pepper as a physiotherapy coach

- open ended wrist turn exercise
- n. reps = *useful* measure of persuasiveness

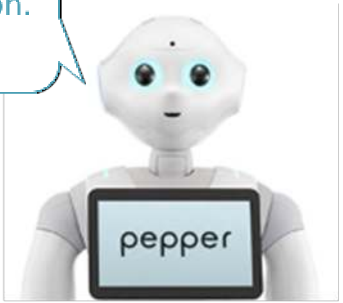



Wizard Station
Participant measures station
X
Participant standing location for interaction

# An Assistive (Persuasive) Social Robot

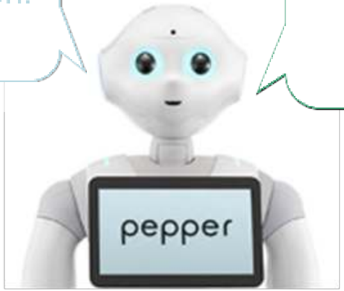
**Expertise**

Tennis elbow is a muscoskeletal condition.
It usually affects…



# An Assistive (Persuasive) Social Robot

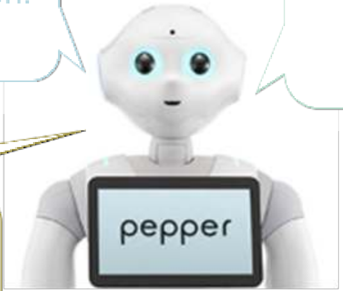
**Expertise**
Tennis elbow is a muscoskeletal condition.
It usually affects...



**Similarity**
Lets compare our preferences. Is it better
to exercise alone or with others?
I also chose....

# An Assistive (Persuasive) Social Robot

**Expertise**
Tennis elbow is a muscoskeletal condition.
It usually affects...



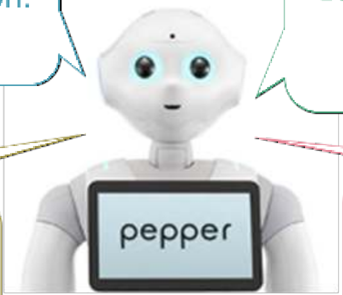
**Similarity**
Lets compare our preferences. Is it better
to exercise alone or with others?
I also chose....

**Goodwill**
How do you feel about today?
Great, I'm glad to hear that! I'm sure
you will enjoy the session.

# An Assistive (Persuasive) Social Robot

**Expertise**
Tennis elbow is a muscoskeletal condition.
It usually affects…

**Similarity**
Lets compare our preferences. Is it better to exercise alone or with others?
I also chose….



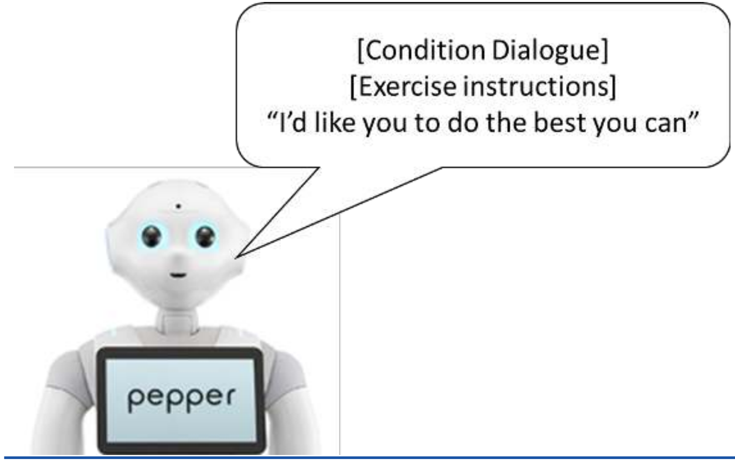
**Goodwill**
How do you feel about today?
Great, I'm glad to hear that! I'm sure you will enjoy the session.

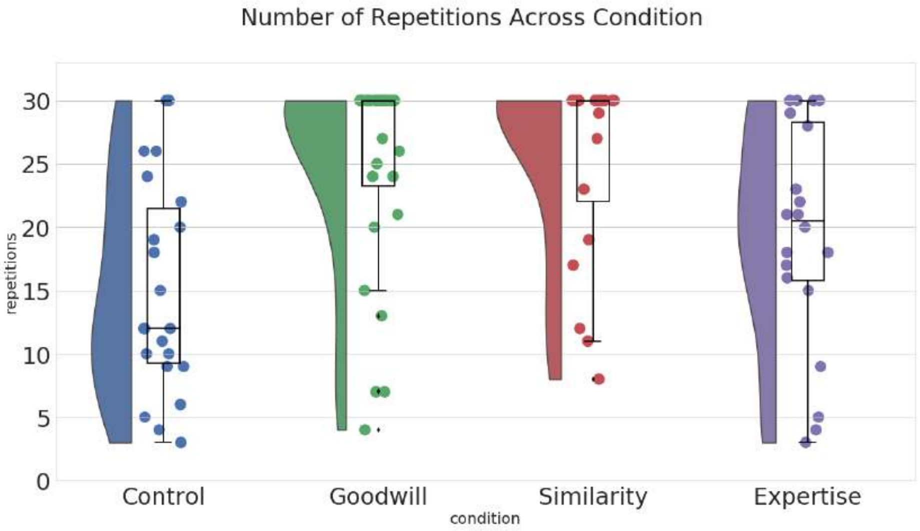
**Control**
I was designed and built by… How did you get here today?

# An Assistive (Persuasive) Social Robot

Pepper as a physiotherapy coach

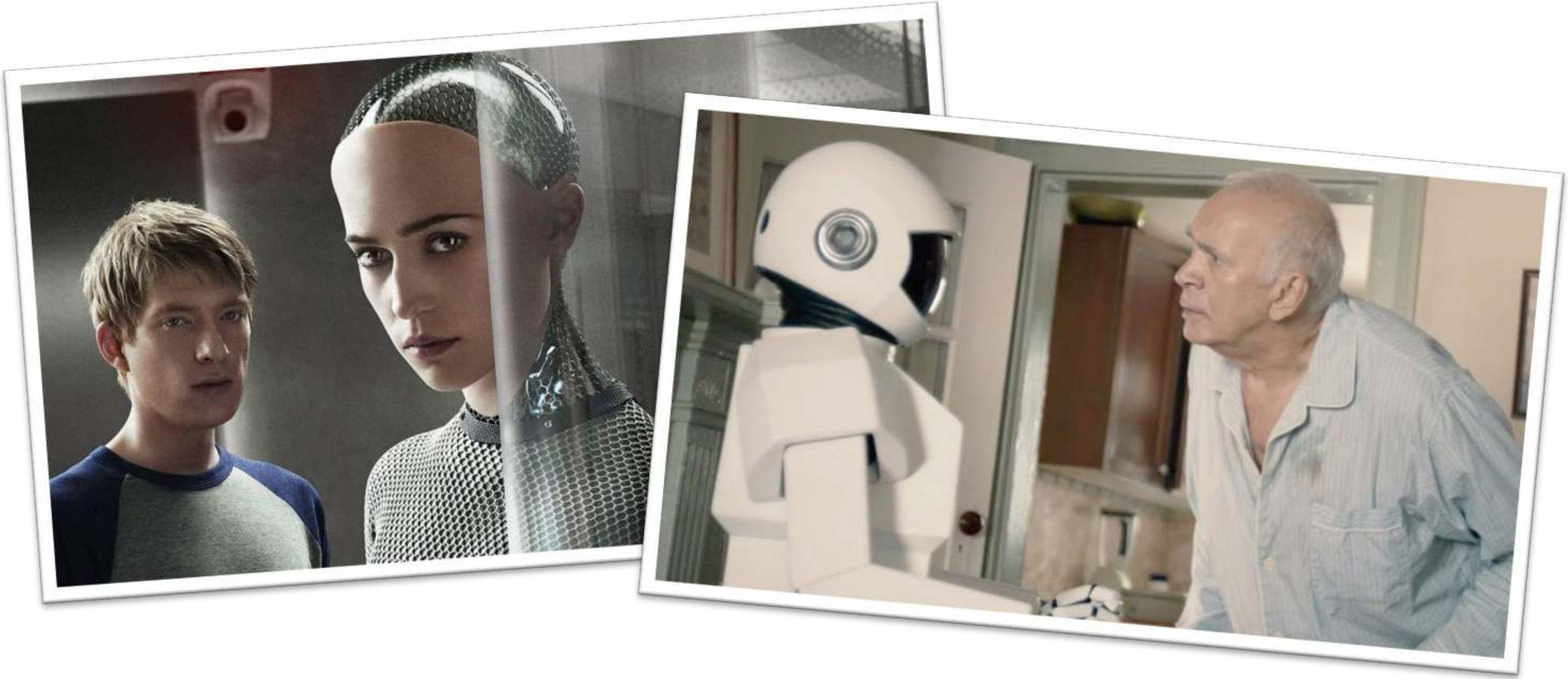
- open ended wrist turn exercise
- n. reps = *useful* measure of persuasiveness





# The Ethical Risk of Artificial Social Behaviour

# The Ethical Risk of Artificial Social Behaviour



# The Ethical Risk of Artificial Social Behaviour

BS 8611:2016

BSI Standards Publication

**Robots and robotic devices**

Guide to the ethical design and application of robots and robotic systems

...making excellence a habit.™

bsi.

# The Ethical Risk of Artificial Social Behaviour

BS 8611:2016

BSI Standards Publication

**Robots and robotic devices**

Guide to the ethical design and application of robots and robotic systems

bsi.

...making excellence a habit.™

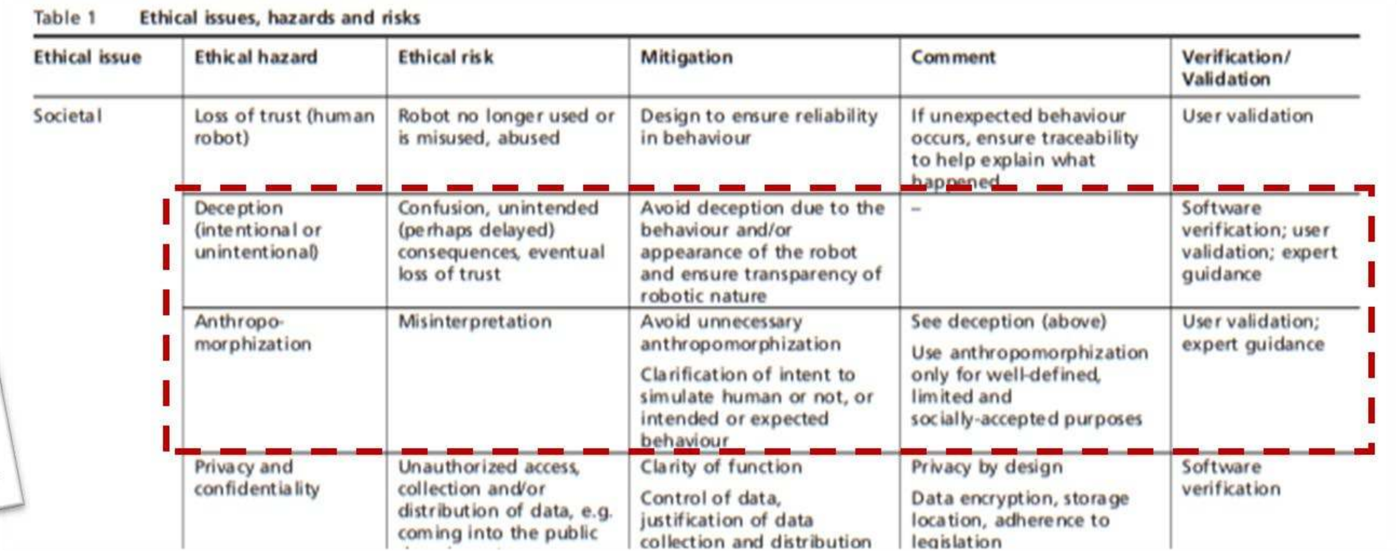


Table 1 **Ethical issues, hazards and risks**

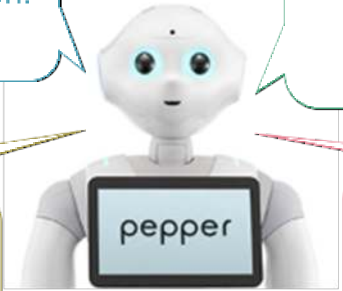
| Ethical issue | Ethical hazard | Ethical risk | Mitigation | Comment | Verification/ Validation |
|---|---|---|---|---|---|
| Societal | Loss of trust (human robot) | Robot no longer used or is misused, abused | Design to ensure reliability in behaviour | If unexpected behaviour occurs, ensure traceability to help explain what happened | User validation |
| | Deception (intentional or unintentional) | Confusion, unintended (perhaps delayed) consequences, eventual loss of trust | Avoid deception due to the behaviour and/or appearance of the robot and ensure transparency of robotic nature | – | Software verification; user validation; expert guidance |
| | Anthropo-morphization | Misinterpretation | Avoid unnecessary anthropomorphization<br>Clarification of intent to simulate human or not, or intended or expected behaviour | See deception (above)<br>Use anthropomorphization only for well-defined, limited and socially-accepted purposes | User validation; expert guidance |
| | Privacy and confidentiality | Unauthorized access, collection and/or distribution of data, e.g. coming into the public | Clarity of function<br>Control of data, justification of data collection and distribution | Privacy by design<br>Data encryption, storage location, adherence to legislation | Software verification |

# The Ethical Risk of Artificial Social Behaviour

(Me, and presumably most other social roboticists…)

# An Assistive, Deceptive Social Robot ?

**Expertise**
Tennis elbow is a muscoskeletal condition.
It usually affects...



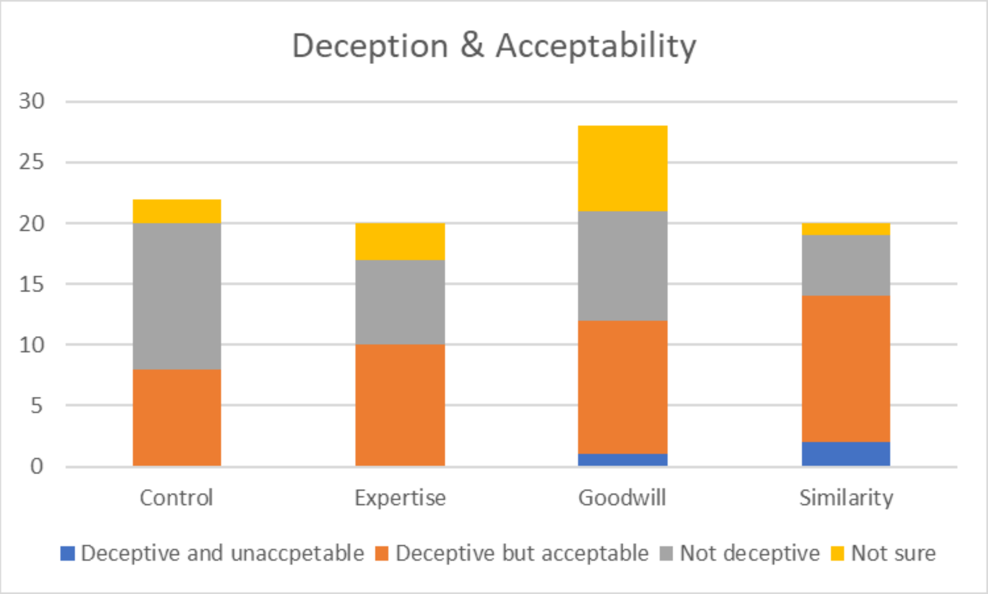
**Similarity**
Lets compare our preferences. Is it better
to exercise alone or with others?
I also chose....

**Goodwill**
How do you feel about today?
Great, I'm glad to hear that! I'm sure
you will enjoy the session.

**Control**
I was designed and built by... How
did you get here today?

# An Assistive, Deceptive Social Robot ?

- Participants generally found the robot *not deceptive* or *deceptive but acceptable*



# An Assistive, Deceptive Social Robot ?

- But this was very much associated with the *application* and the link back to other humans behind the scenes – complex reasoning!

# An Assistive, Deceptive Social Robot ?

- But this was very much associated with the *application* and the link back to other humans behind the scenes – complex reasoning!

> *“I felt like it was genuine but also I'm very aware that somebody else programmed it to be genuine, but I'm ok with that because I feel like* ***whoever had made the programme*** *in the first place did want the person [exercising] to feel comfortable and to feel cared about...it's the intention behind it.”*

# How Much Deception Do We *Need?*

<u>‘Higher Risk’ Social Behaviour</u>

*You’re from Bristol, just like me! I live in the Bristol Robotics Lab.*

*I know that exercising can be boring and hard, and we all suffer from a lack of motivation sometimes. I hope I can make exercising a bit more enjoyable for you.*

*That was great, I’m very impressed.*

<u>‘Lower Risk’ Social Behaviour</u>

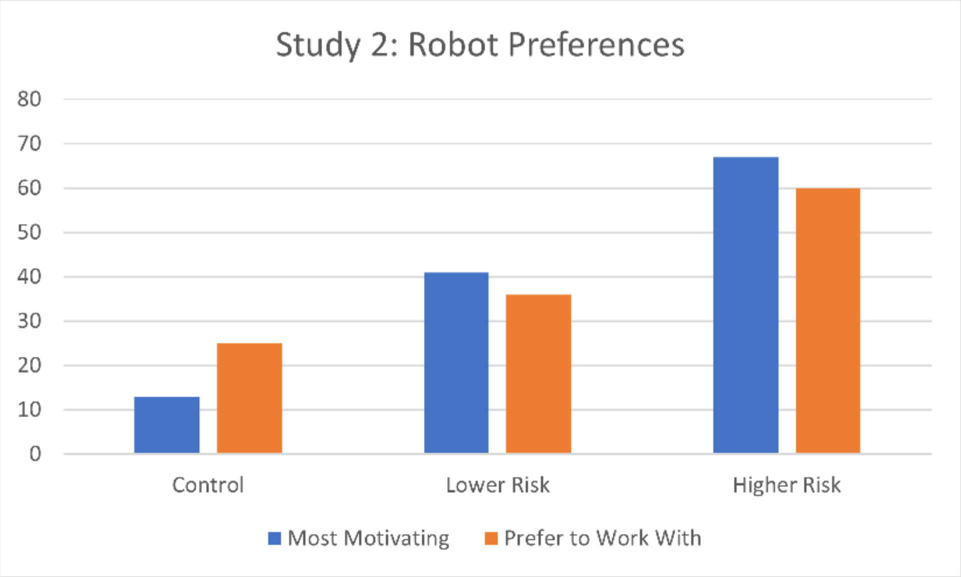
*The robotics lab where I was programmed is also in Bristol.*

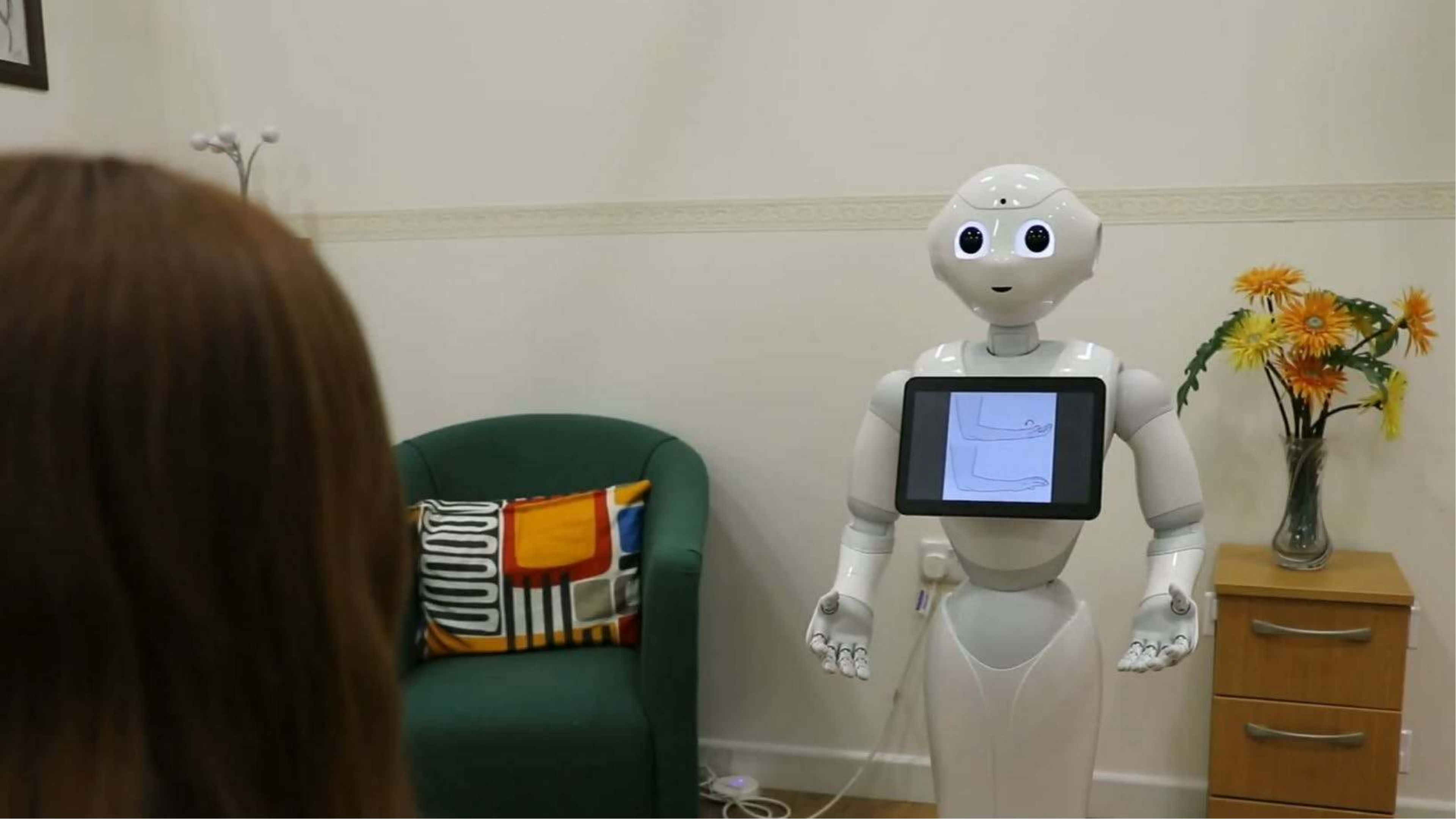
*Many patients find exercising boring or hard, and it is normal to suffer from a lack of motivation sometimes. Perhaps working with me will make exercising a bit more enjoyable for you.*

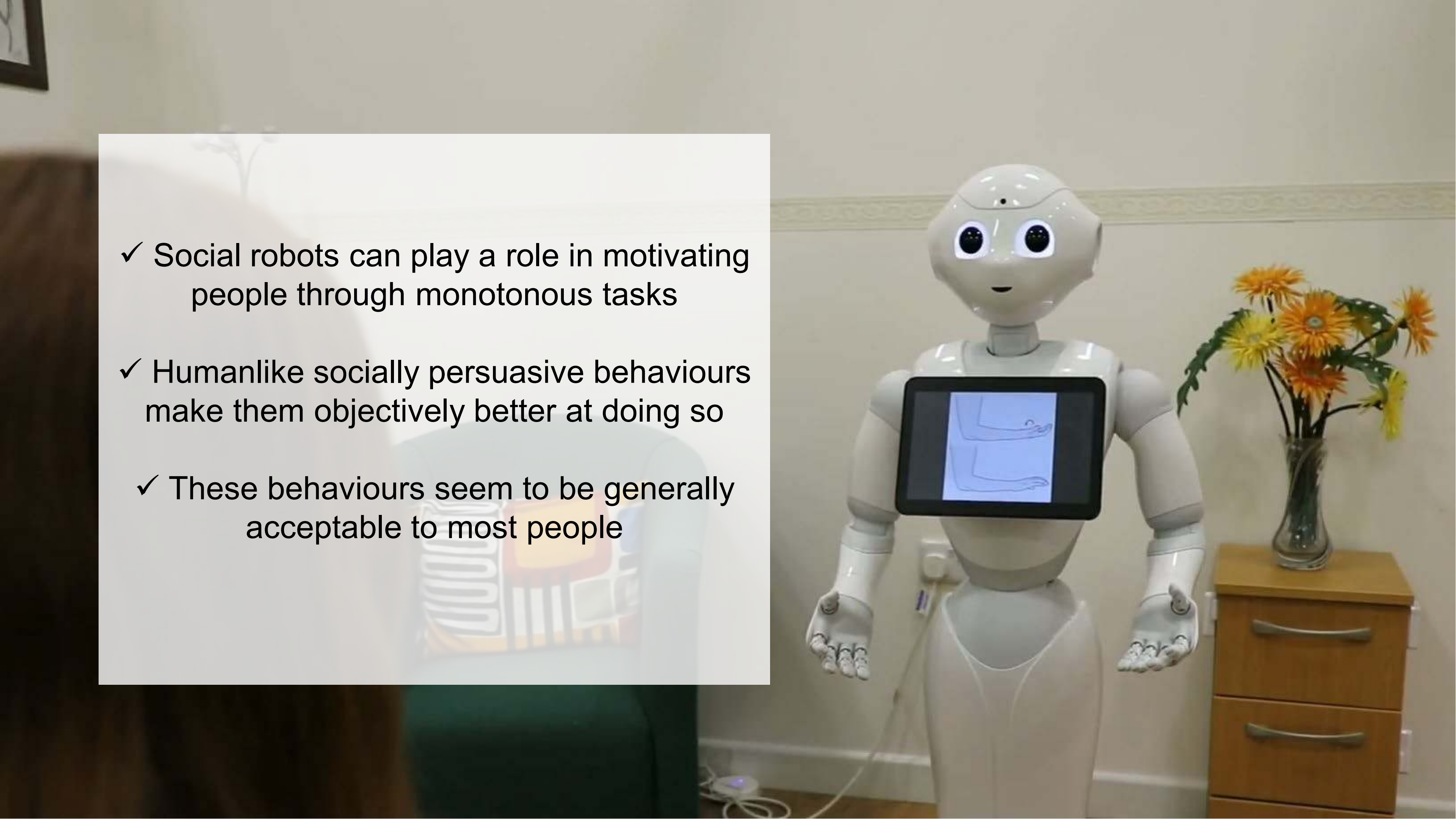
*That was good, your therapist would be impressed.*

# How Much Deception Do We *Need?*

- Higher risk robot had greater credibility than lower risk and control robots
- Higher risk robot most preferred







✓ Social robots can play a role in motivating people through monotonous tasks

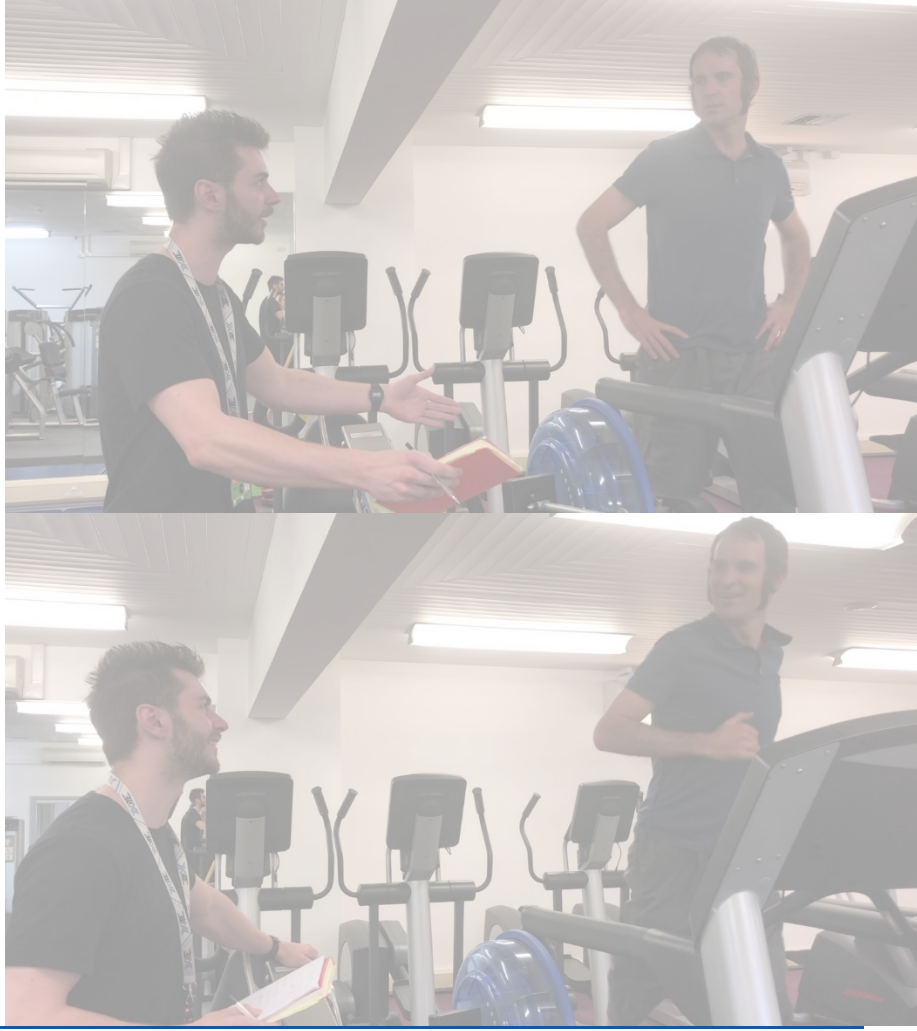
✓ Humanlike socially persuasive behaviours make them objectively better at doing so

✓ These behaviours seem to be generally acceptable to most people

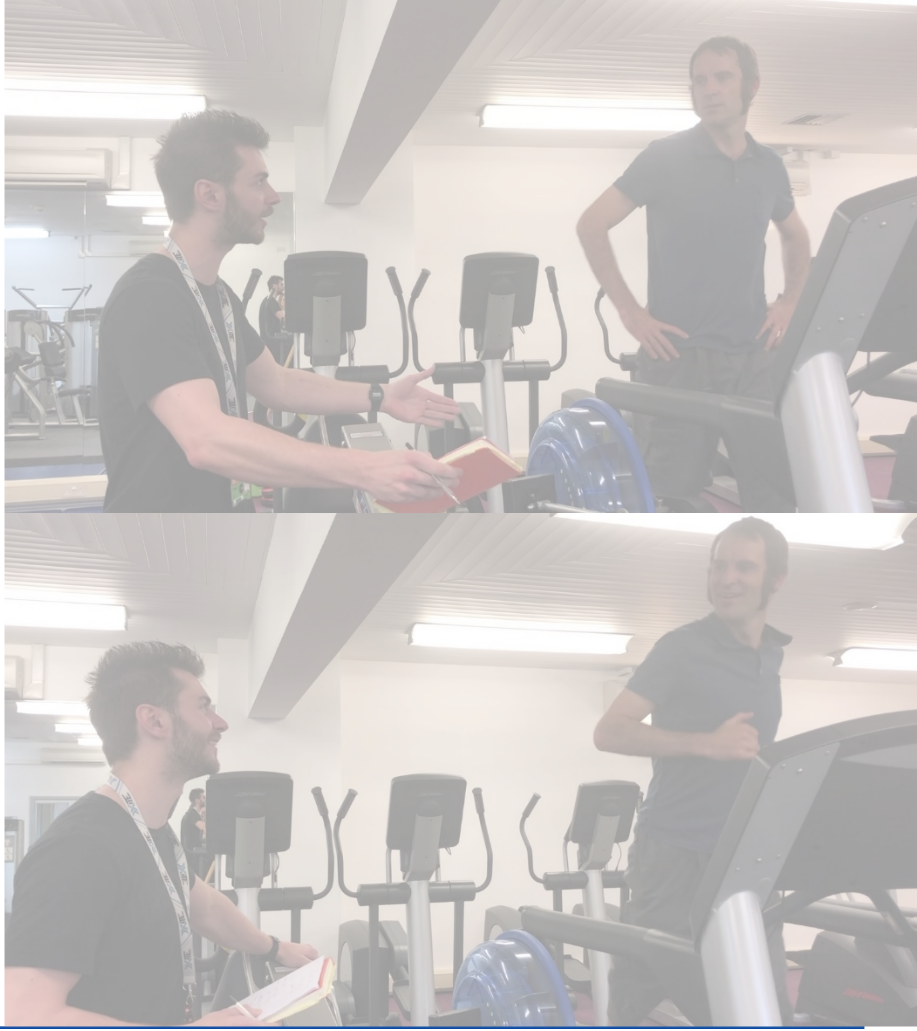
(Human) experts in this know when (and how) to be more serious and informative…

…but they also know when (and how) to be more fun; and how to do that differently across different clients to build good rapport and keep them engaged.

How do we go about designing and automating such complex, tacit, intangible social intelligence?!



How could / should a robot fit into this picture?



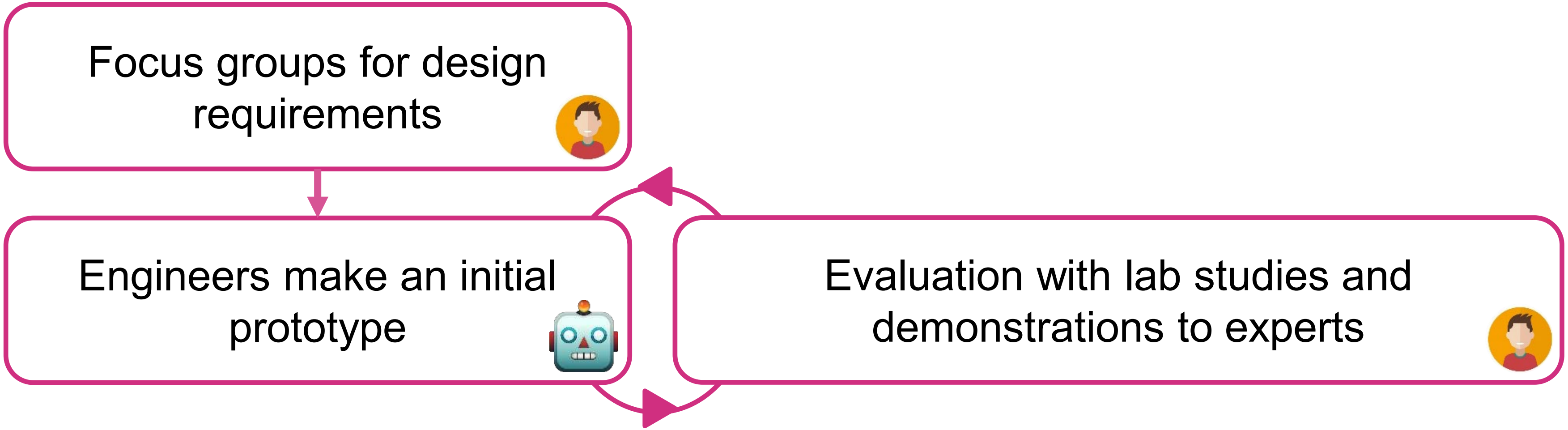
# Working with humans to make robots better

It seems *obvious* that we should be working with domain experts (and other stakeholders) in designing socially assistive robots.

# Participatory Social Robot Design

Focus groups for design requirements

# Participatory Social Robot Design

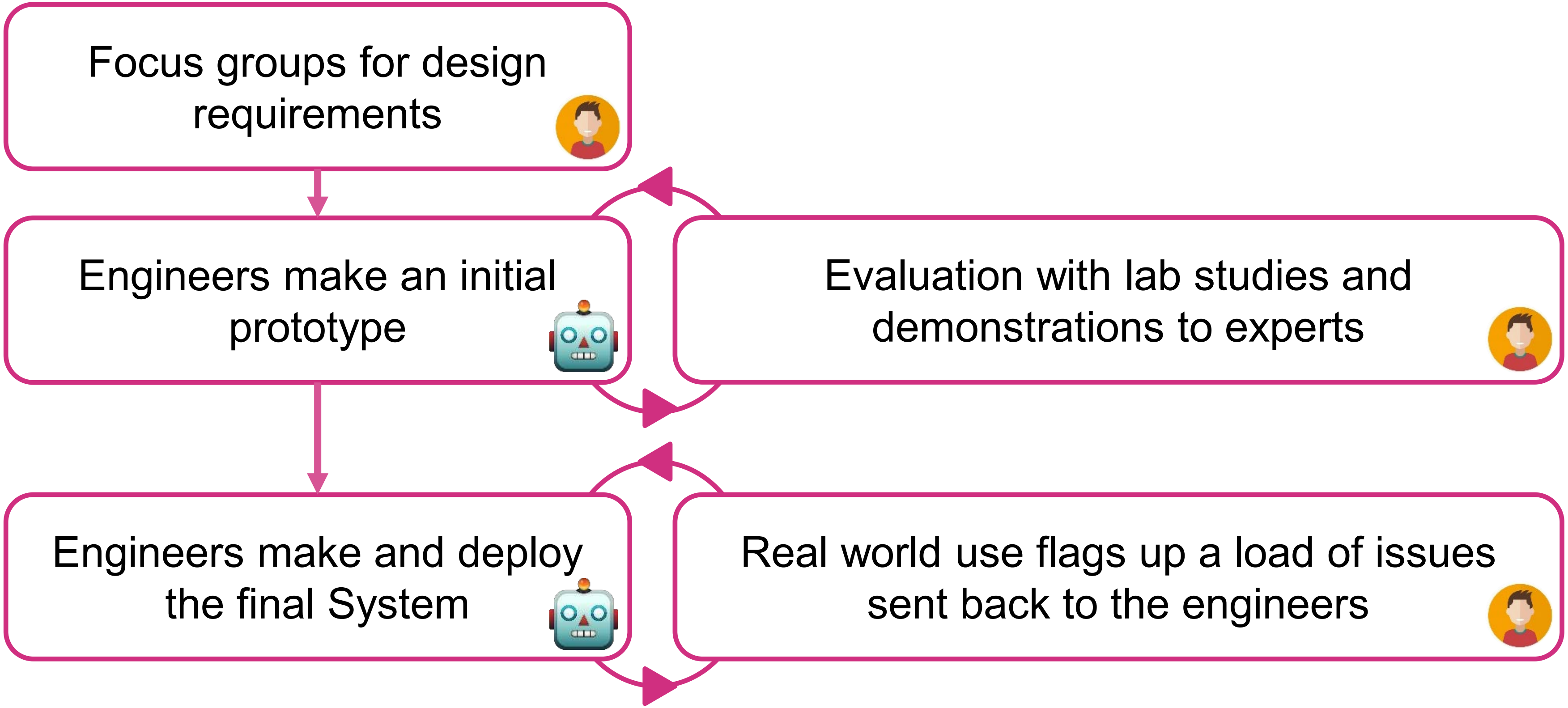


Focus groups for design requirements

Engineers make an initial prototype

Evaluation with lab studies and demonstrations to experts

# Participatory Social Robot Design



Focus groups for design requirements

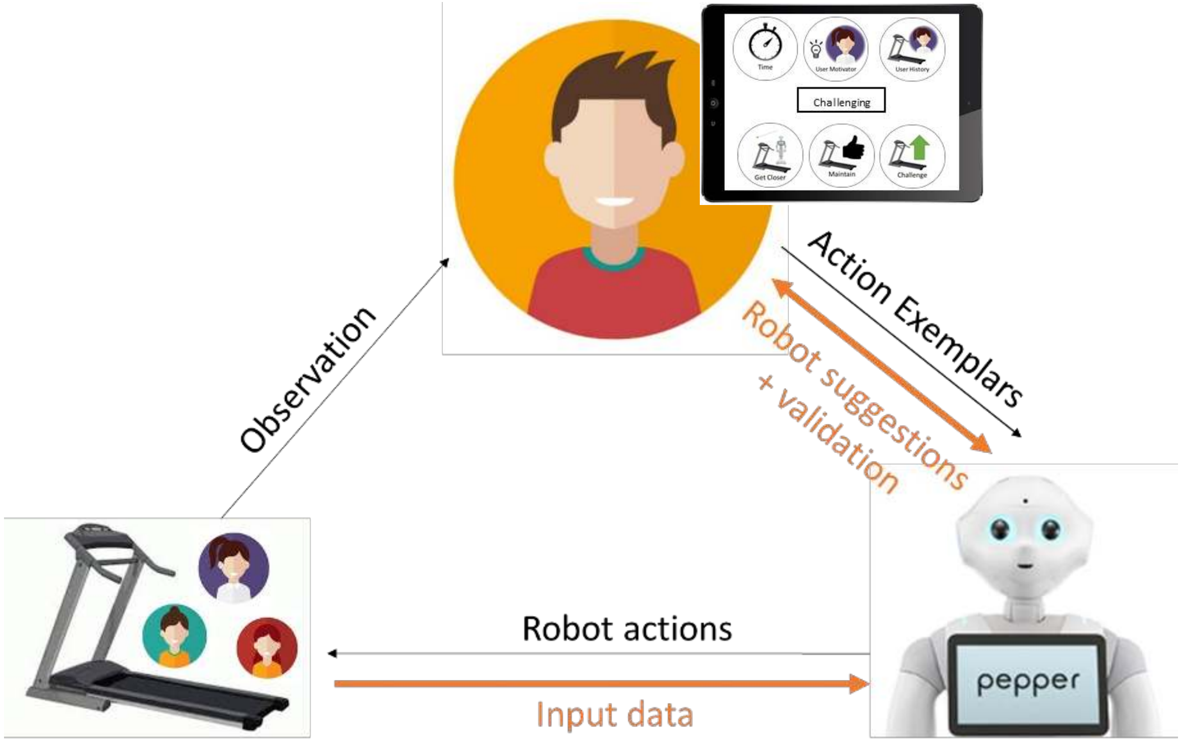
Engineers make an initial prototype

Evaluation with lab studies and demonstrations to experts

Engineers make and deploy the final System

Real world use flags up a load of issues sent back to the engineers

# Expert-in-the-Loop Interactive Machine Learning…



Observation

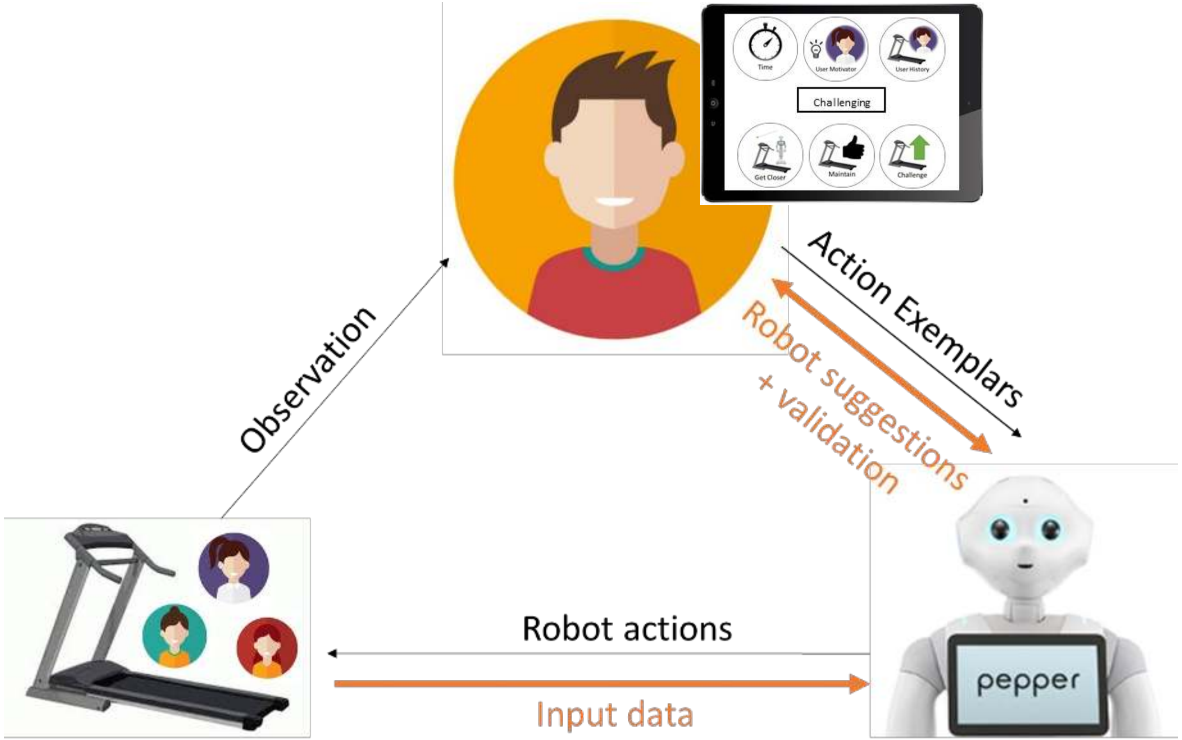
Action Exemplars

Robot suggestions + validation

Robot actions

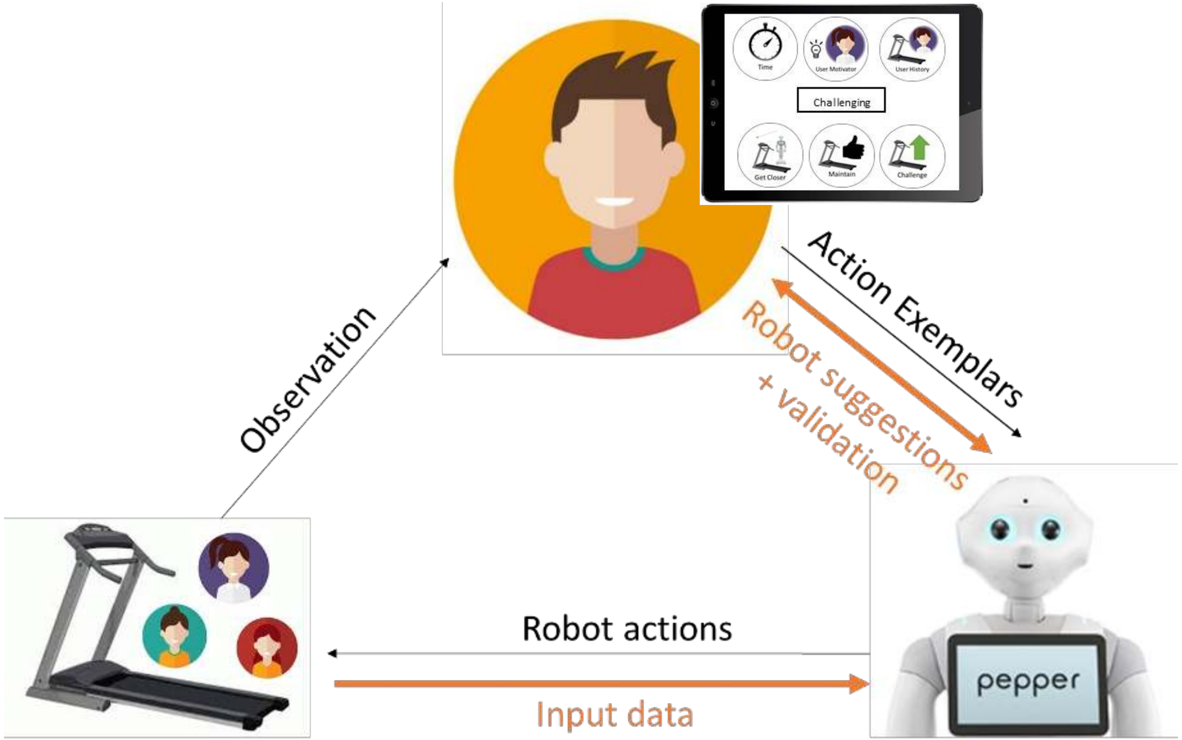
Input data

# Expert-in-the-Loop Interactive Machine Learning…



1. Co-design robot actions and input space with a *domain* expert

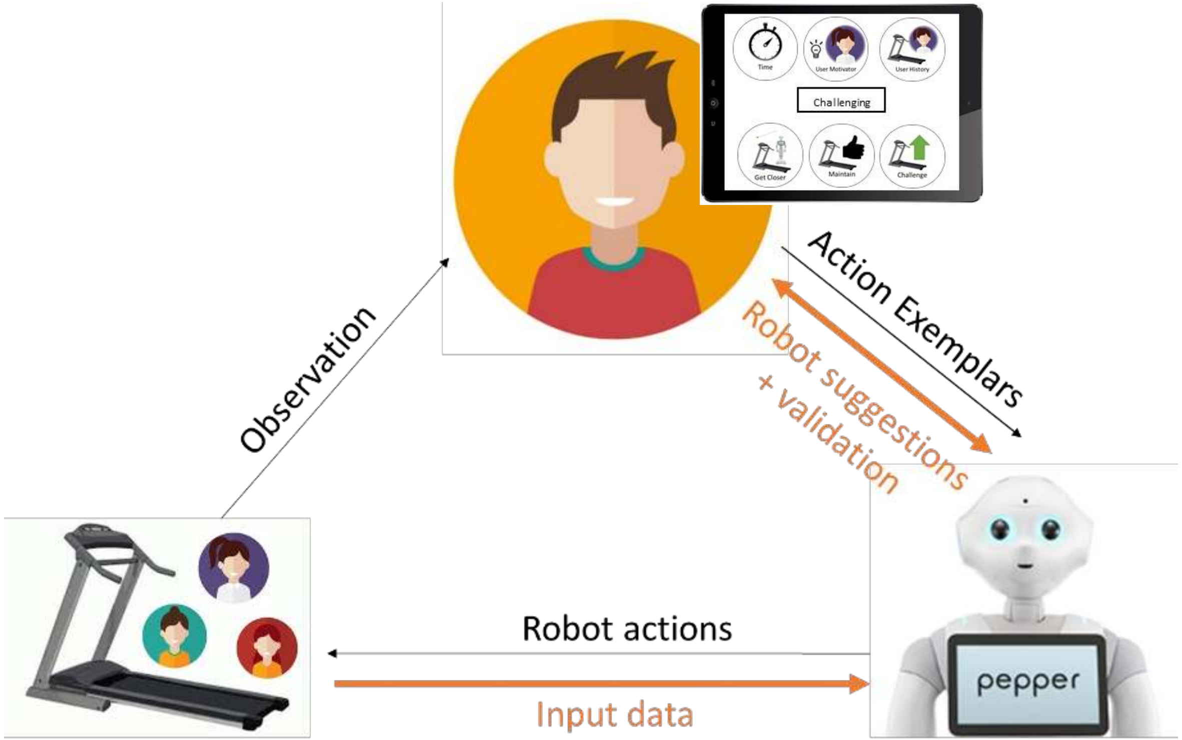
# Expert-in-the-Loop Interactive Machine Learning…



1. Co-design robot actions and input space with a *domain* expert

2. Co-design a ‘teaching interface’ for using those actions and responding to robot suggestions

# Expert-in-the-Loop Interactive Machine Learning…

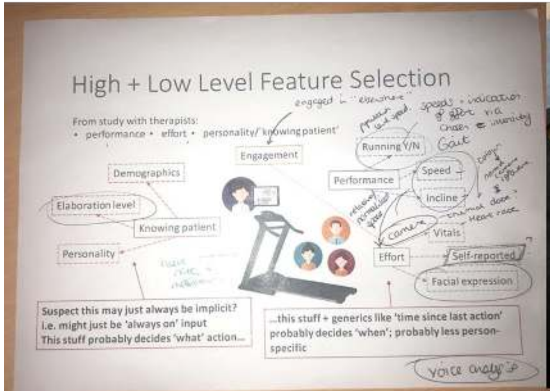


1. Co-design robot actions and input space with a *domain* expert

2. Co-design a ‘teaching interface’ for using those actions and responding to robot suggestions

3. Domain expert *teaches* the robot via interactive machine learning in-the-wild

# … as Participatory Design



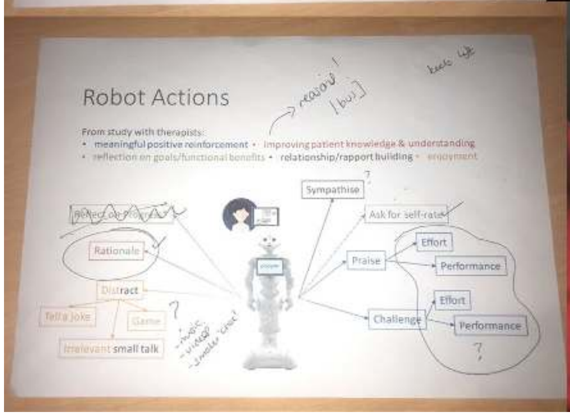
High + Low Level Feature Selection

From study with therapists:
- performance • effort • personality/'knowing patient'

Suspect this may just always be implicit? i.e. might just be 'always on' input This stuff probably decides 'what' action...

...this stuff + generics like 'time since last action' probably decides 'when'; probably less person-specific

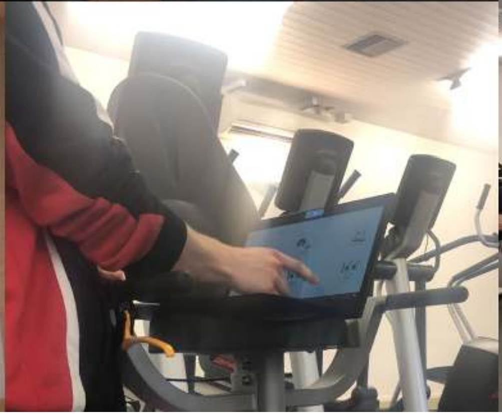
Engagement
Demographics
Elaboration level
Knowing patient
Personality
Running Y/N
Performance
Speed
Incline
Vitals
Effort
Self-reported
Facial expression
engaged ↳ "elsewhere"
speeds ↳ indication of effort via chosen intensity
Gait
camera
Heart rate
voice analysis



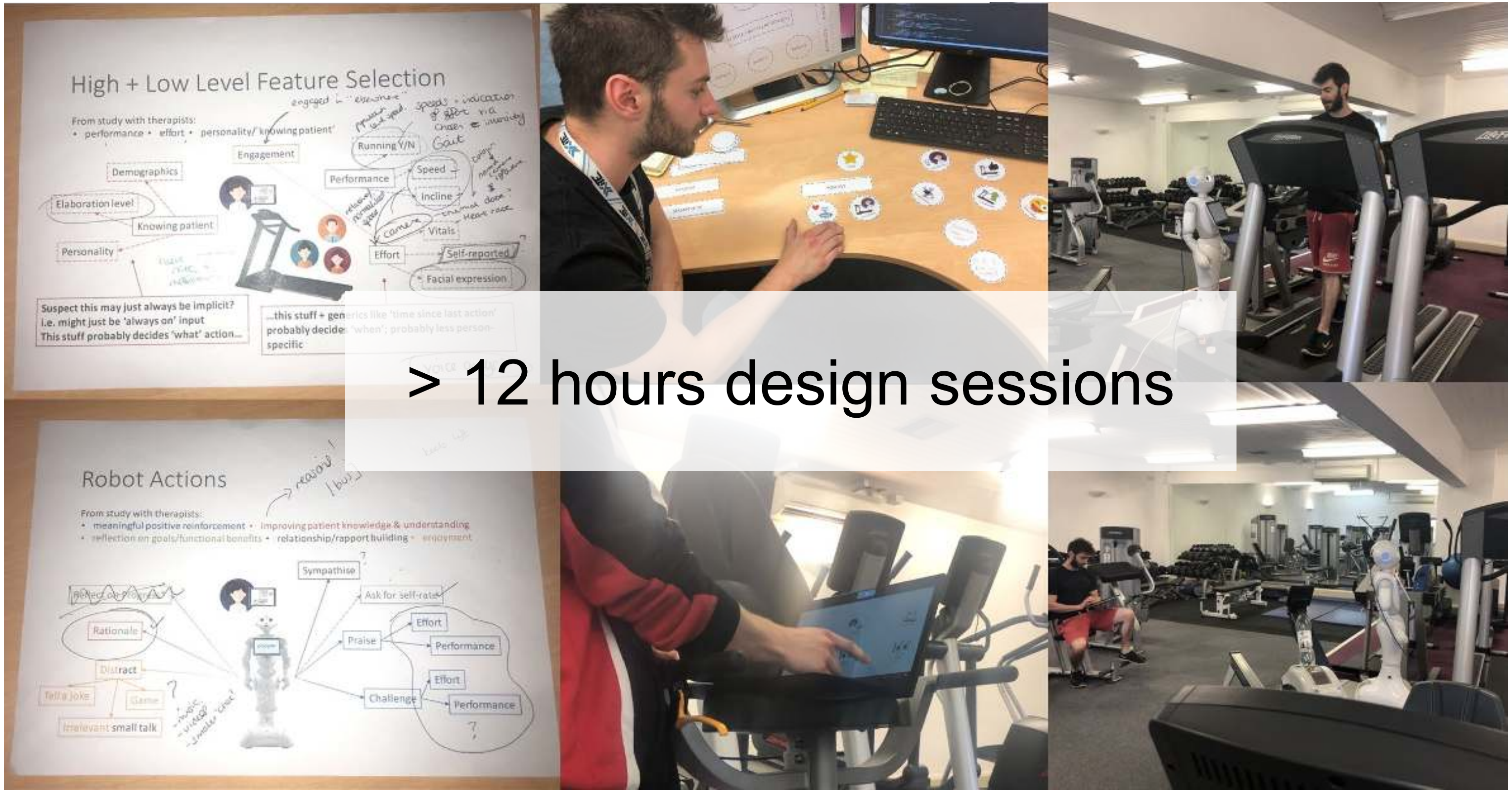
Robot Actions

From study with therapists:
- meaningful positive reinforcement • improving patient knowledge & understanding
- reflection on goals/functional benefits • relationship/rapport building • enjoyment

reason! [but]
Sympathise
~~Reflect on Progress~~
Ask for self-rate
Rationale
Praise
Effort
Performance
Distract
Tell a Joke
Game
Irrelevant small talk
Challenge
Effort
Performance
- music
- video
- smaller chat
?



# … as Participatory Design

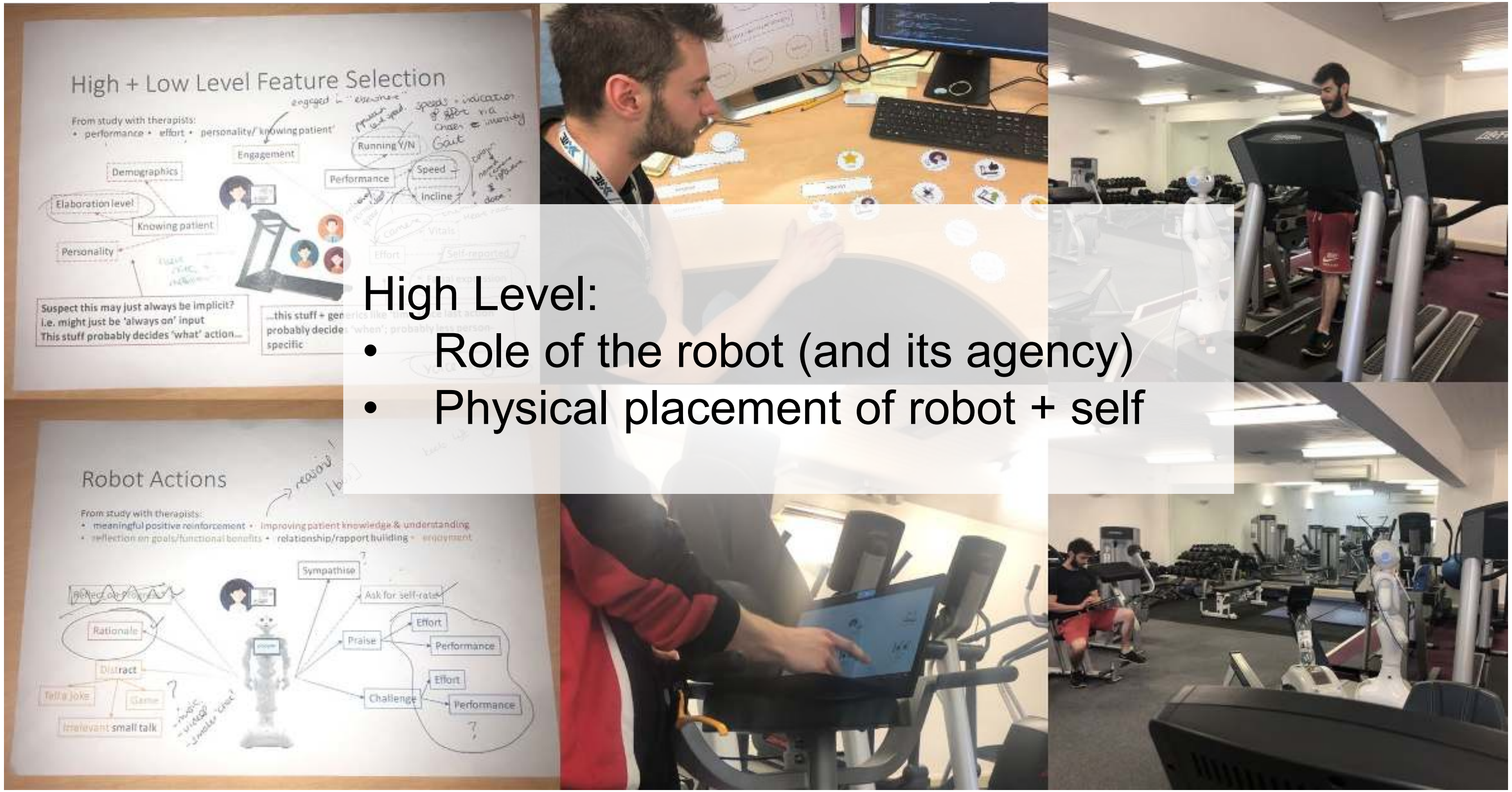


Suspect this may just always be implicit?
i.e. might just be 'always on' input
This stuff probably decides 'what' action…

…this stuff + generics like 'time since last action' probably decides 'when'; probably less person-specific

> 12 hours design sessions

# … as Participatory Design



High + Low Level Feature Selection

From study with therapists:
- performance • effort • personality/'knowing patient'

Suspect this may just always be implicit? i.e. might just be 'always on' input This stuff probably decides 'what' action...

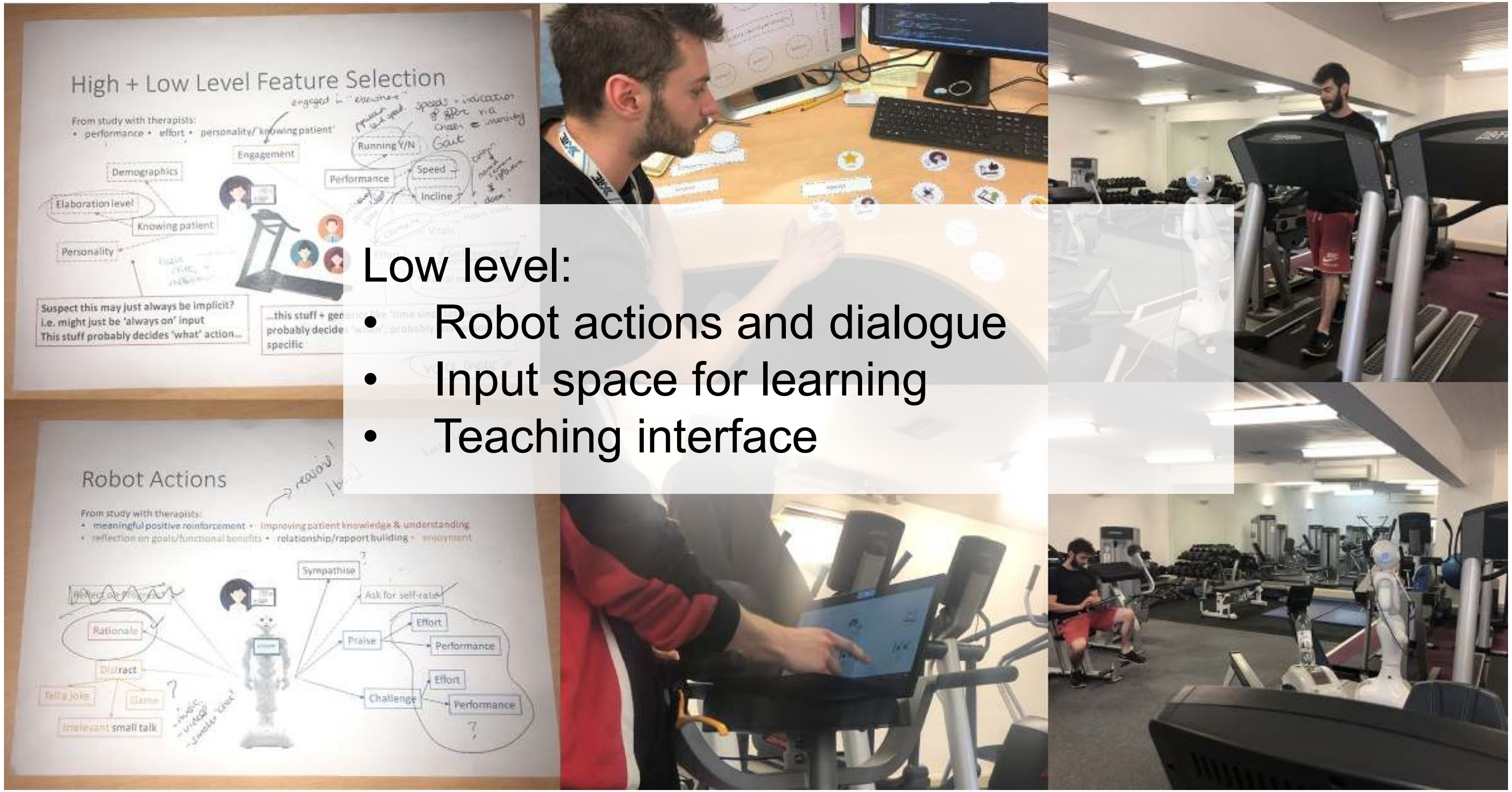
...this stuff + generics like 'time since last action' probably decides 'when'; probably less person-specific

Robot Actions

From study with therapists:
- meaningful positive reinforcement • improving patient knowledge & understanding
- reflection on goals/functional benefits • relationship/rapport building • enjoyment

High Level:
- Role of the robot (and its agency)
- Physical placement of robot + self

# … as Participatory Design



High + Low Level Feature Selection

From study with therapists:
- performance
- effort
- personality/ 'knowing patient'

Suspect this may just always be implicit?
i.e. might just be 'always on' input
This stuff probably decides 'what' action…

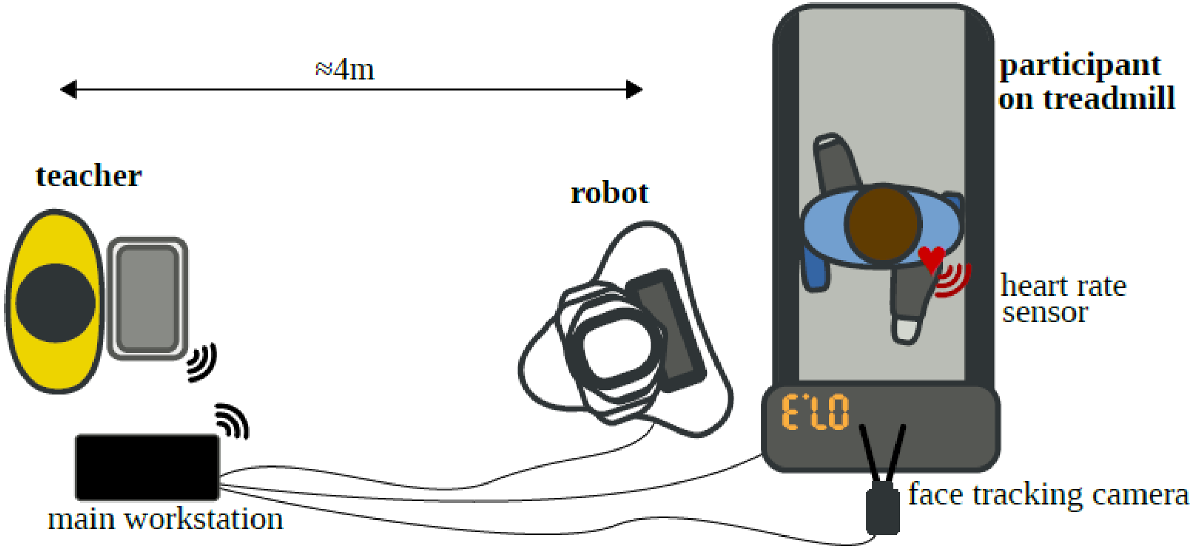
…this stuff + generics like 'time since last action'
probably decides 'when'; probably less person specific

Robot Actions

From study with therapists:
- meaningful positive reinforcement
- improving patient knowledge & understanding
- reflection on goals/functional benefits
- relationship/rapport building
- enjoyment

Low level:
- Robot actions and dialogue
- Input space for learning
- Teaching interface

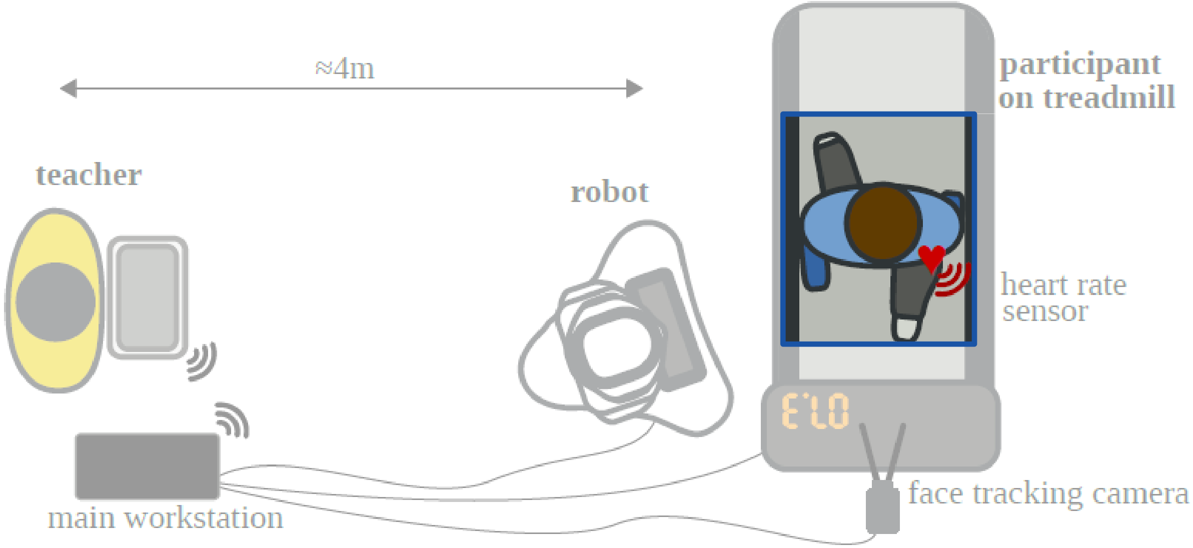
# … as Participatory Design



Input space covered the ‘typical’:

- task state
- performance (*speed*)
- effort (*heart, face*)

# … as Participatory Design



but also:

- overall ‘motivation’
- personality

# … as Participatory Design

“Can you push a bit harder? Maybe turn up the speed” ***speed up***

“I’m impressed, you’re doing great!” ***praise***

“You can call me terminator because I’m going to make you run for your life” ***humour***

# … for Successful HRI in the Real World

✓ Installed the robot in an actual university gym for 3 months and delivered a functional exercise program to 10 participants – of whom only 1 dropped out!

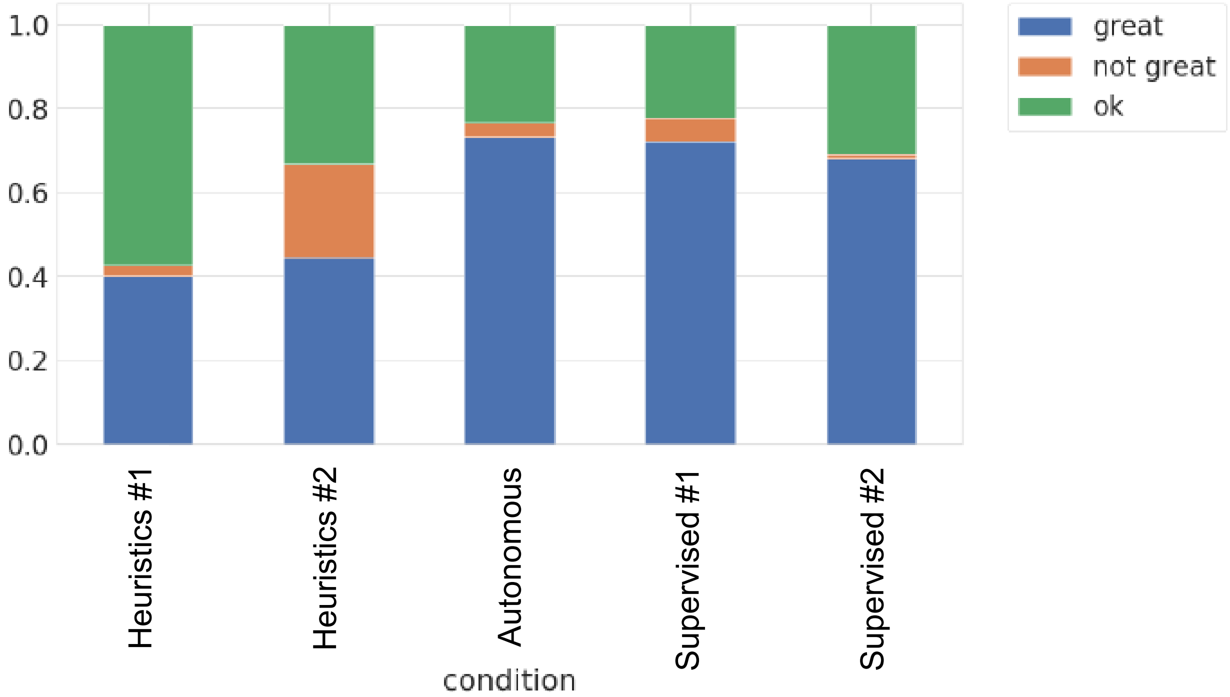
# … for Successful HRI in the Real World

- ✓ Installed the robot in an actual university gym for 3 months and delivered a functional exercise program to 10 participants – of whom only 1 dropped out!

- ✓ Ran a total of **232 robot-led, instructor-supported sessions**:
  - ✓ 151 supervised sessions (for training data)
  - ✓ 32 autonomous with the IML trained system
  - ✓ 49 autonomous through heuristics (a ‘control’ condition)

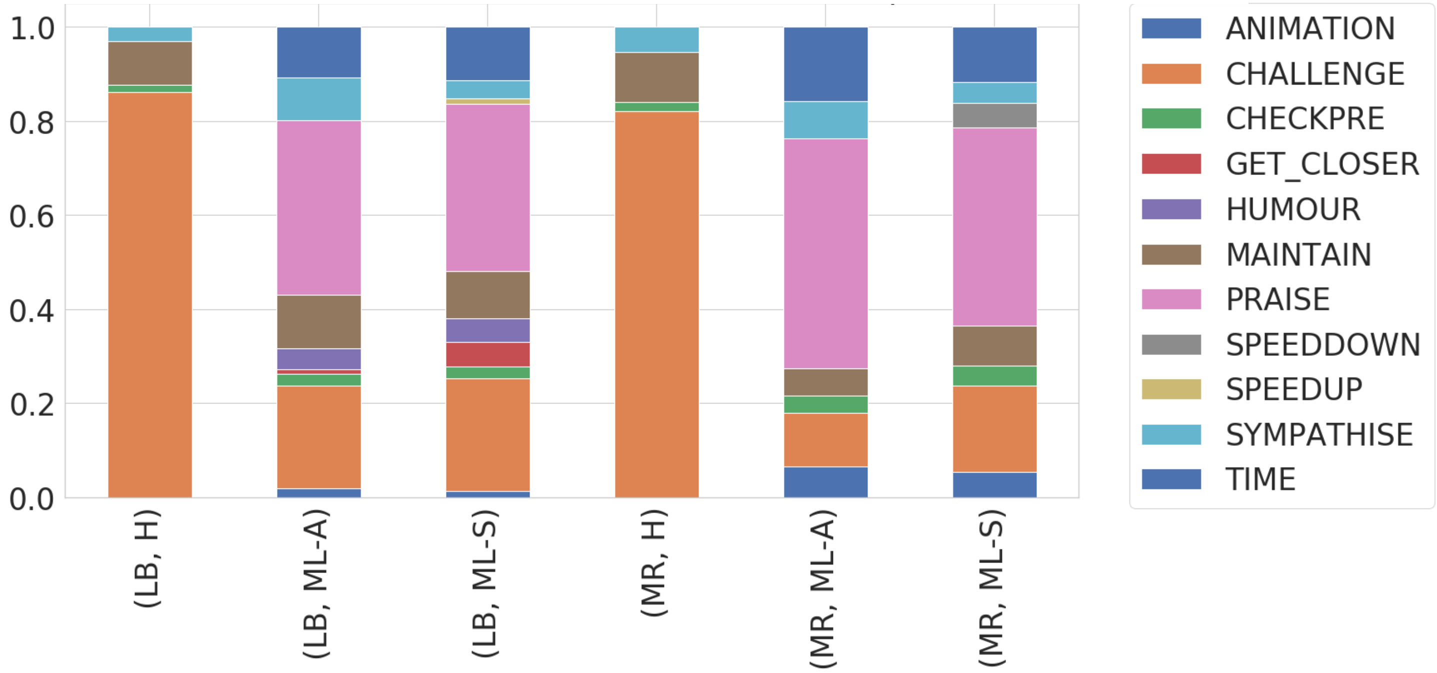
# … for (Good!) Autonomous Robot Behaviour

Post-Session Evaluation Scores for *Heuristic, Autonomous* + *Supervised* Sessions



# … for (Good!) Autonomous Robot Behaviour

*Heuristic, Autonomous* + *Supervised* Action Distribution for Two Different Participants



# … for Successful HRI in the Real World

**Good Autonomous Behaviour**

Autonomous robot learned appropriate action policy.

Was not rated significantly different to supervise system… only 2 participants noticed the switch!

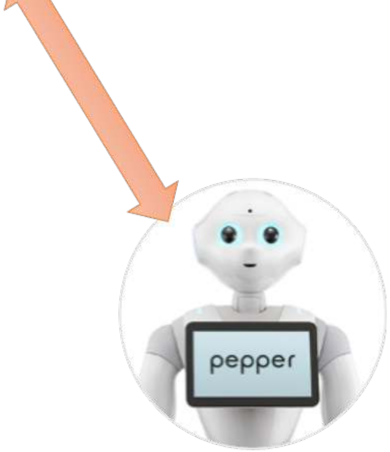
# … for Successful HRI in the Real World

### Emergent Synergy

Unplanned: instructor used robot to autonomously lead warm ups while he did some post-run stretches with the previous participant.

# … for Successful HRI in the Real World

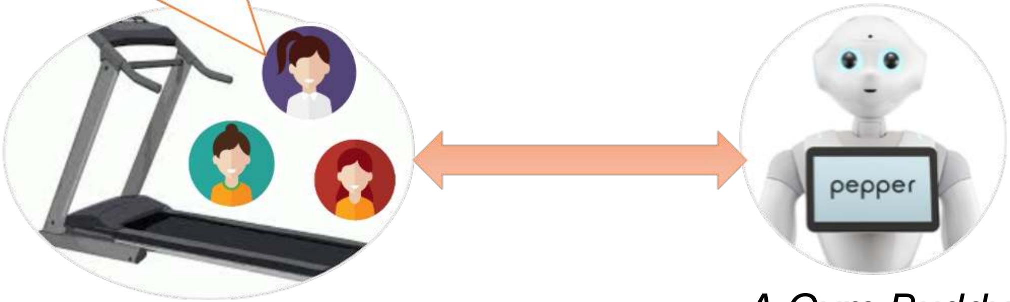
When the sessions got busier… we could work together doing separate things but to get more work done and I think that that's more of a teammate colleague trait than a tool.



*A Colleague*

# … for Successful HRI in the Real World

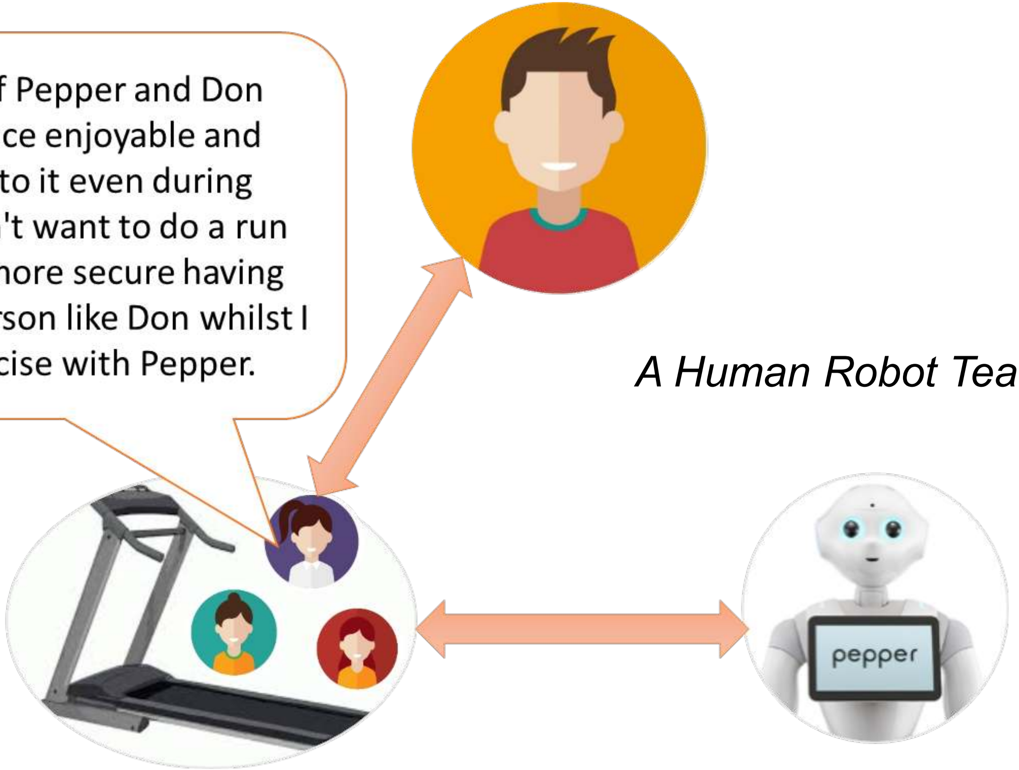
It was a great gym-buddy companion that made me want to go to the session and try my best.



*A Gym Buddy*

# … for Successful HRI in the Real World

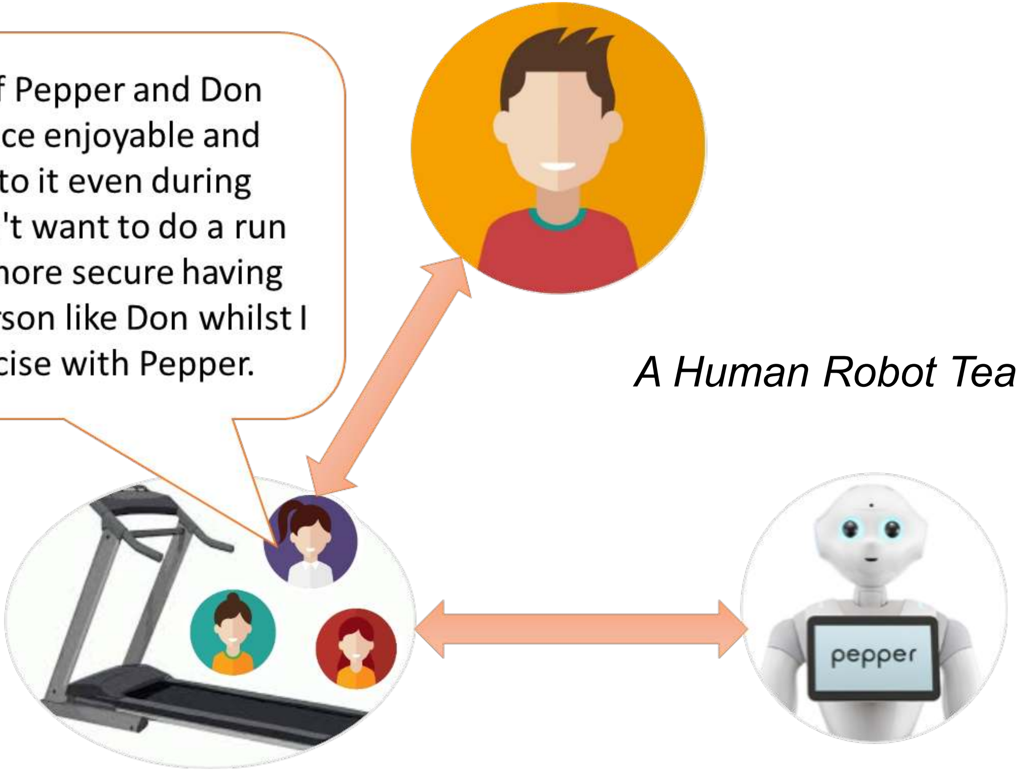
The combination of Pepper and Don made this experience enjoyable and helped me to stick to it even during the days that I didn't want to do a run at all. I think I felt more secure having an experienced person like Don whilst I was doing the exercise with Pepper.



*A Human Robot Team*

# … for Successful HRI in the Real World

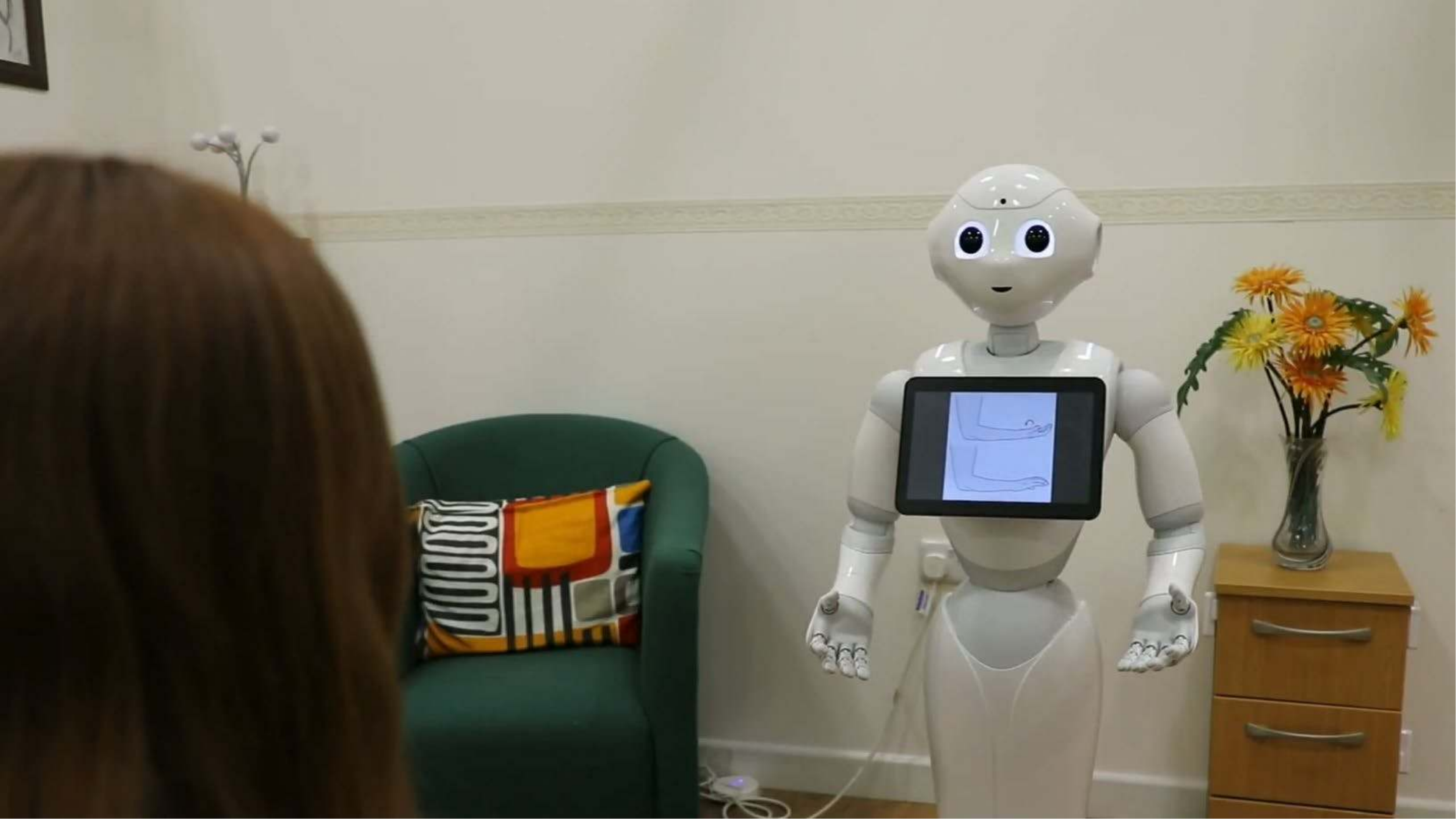
The combination of Pepper and Don made this experience enjoyable and helped me to stick to it even during the days that I didn't want to do a run at all. I think I felt more secure having an experienced person like Don whilst I was doing the exercise with Pepper.

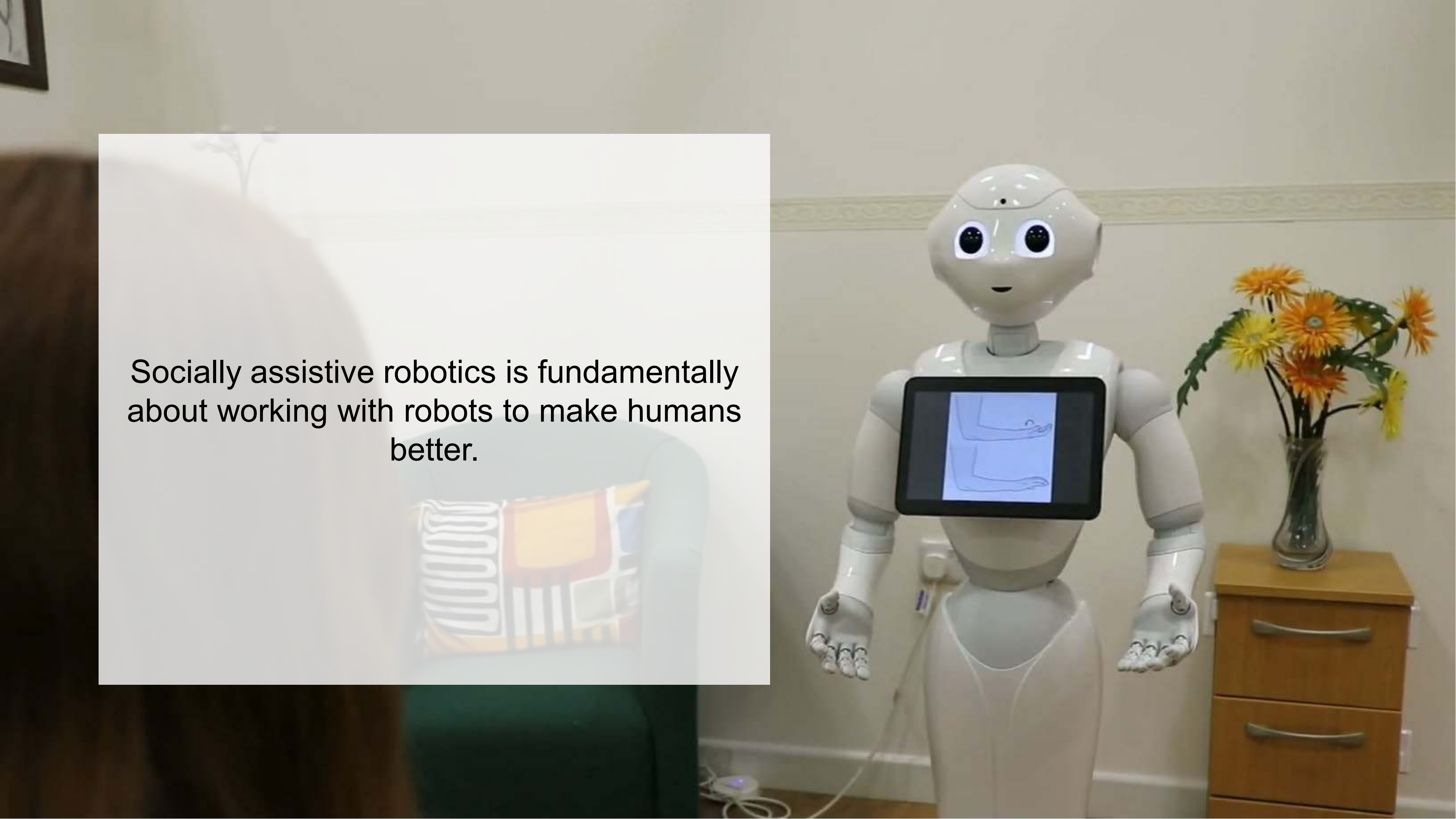


*A Human Robot Team*

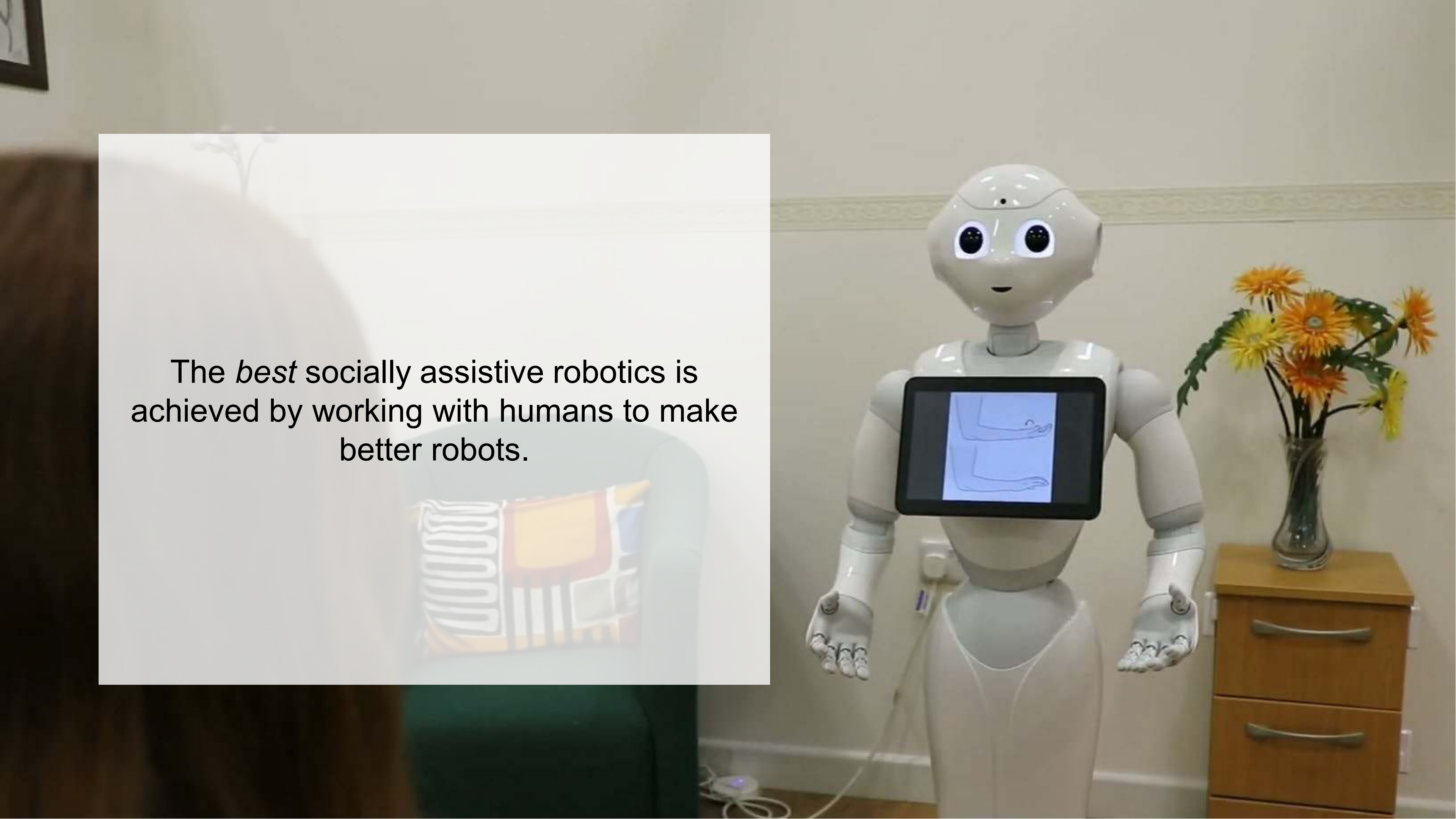
▶

# … for Successful HRI in the Real World



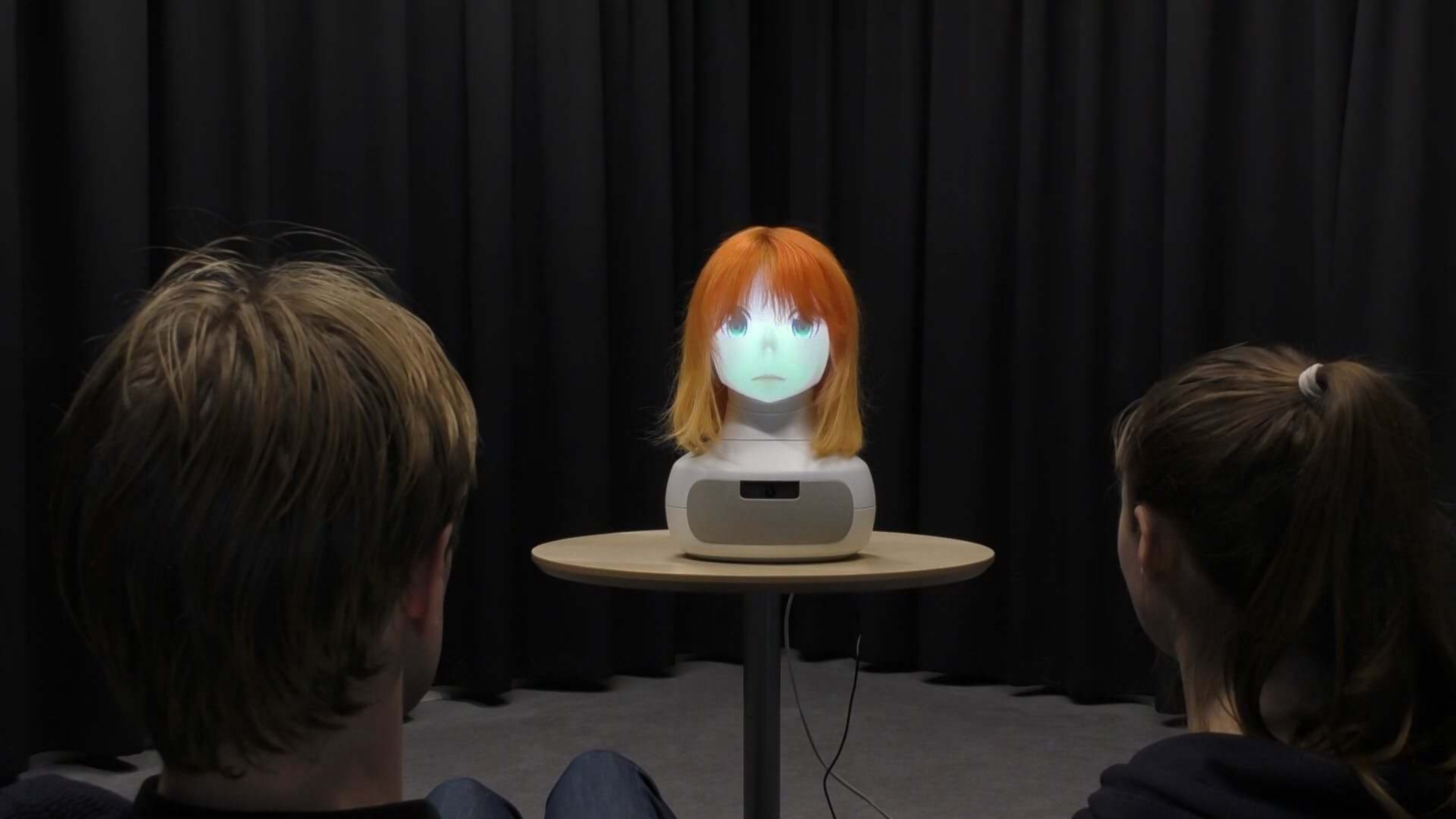


Socially assistive robotics is fundamentally about working with robots to make humans better.



The *best* socially assistive robotics is achieved by working with humans to make better robots.

# Working with robots to make better humans…?



I'd blush if I could

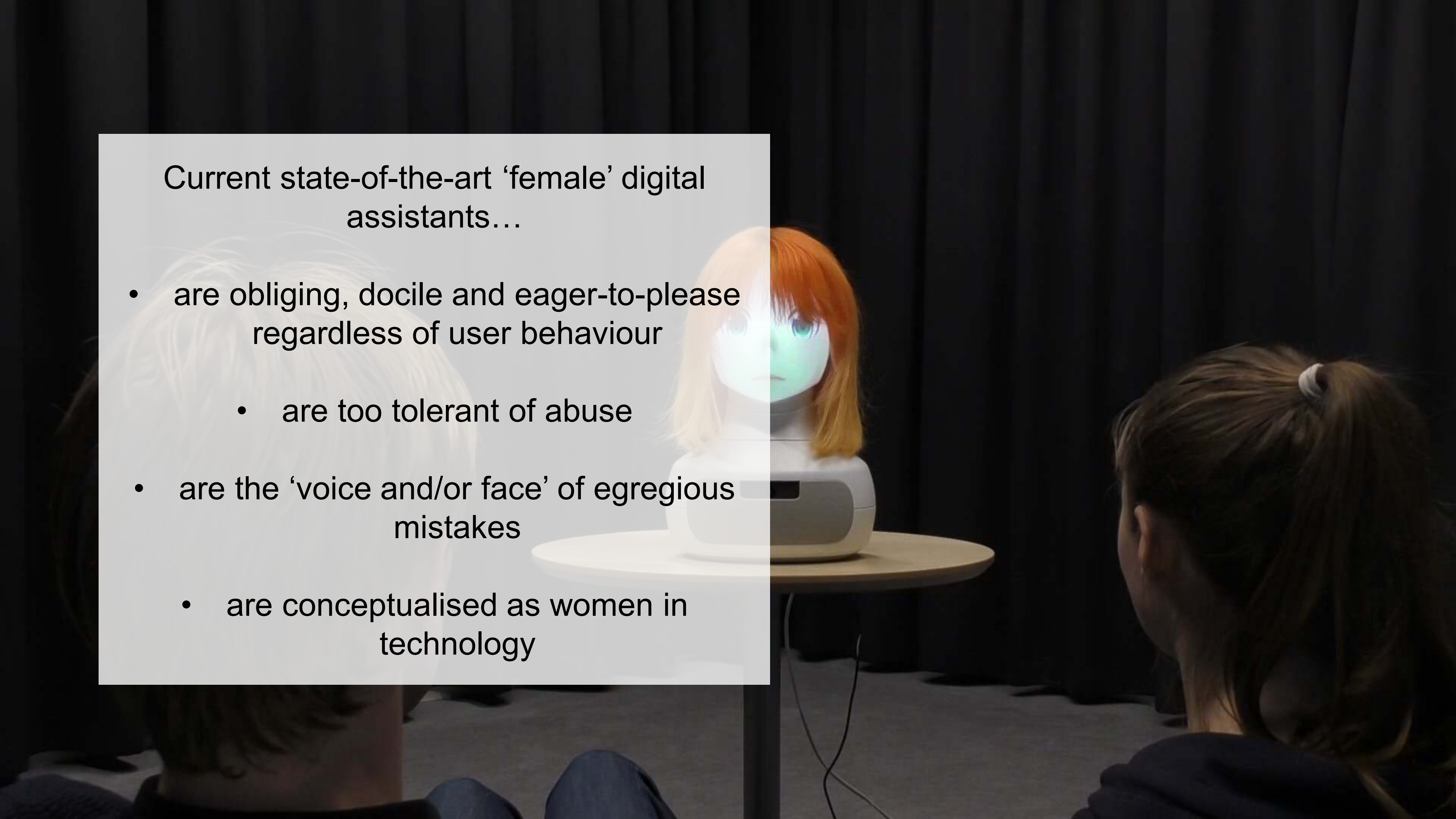
CLOSING GENDER DIVIDES IN DIGITAL SKILLS THROUGH EDUCATION

EQUALS
GLOBAL PARTNERSHIP

United Nations
Educational, Scientific and
Cultural Organization

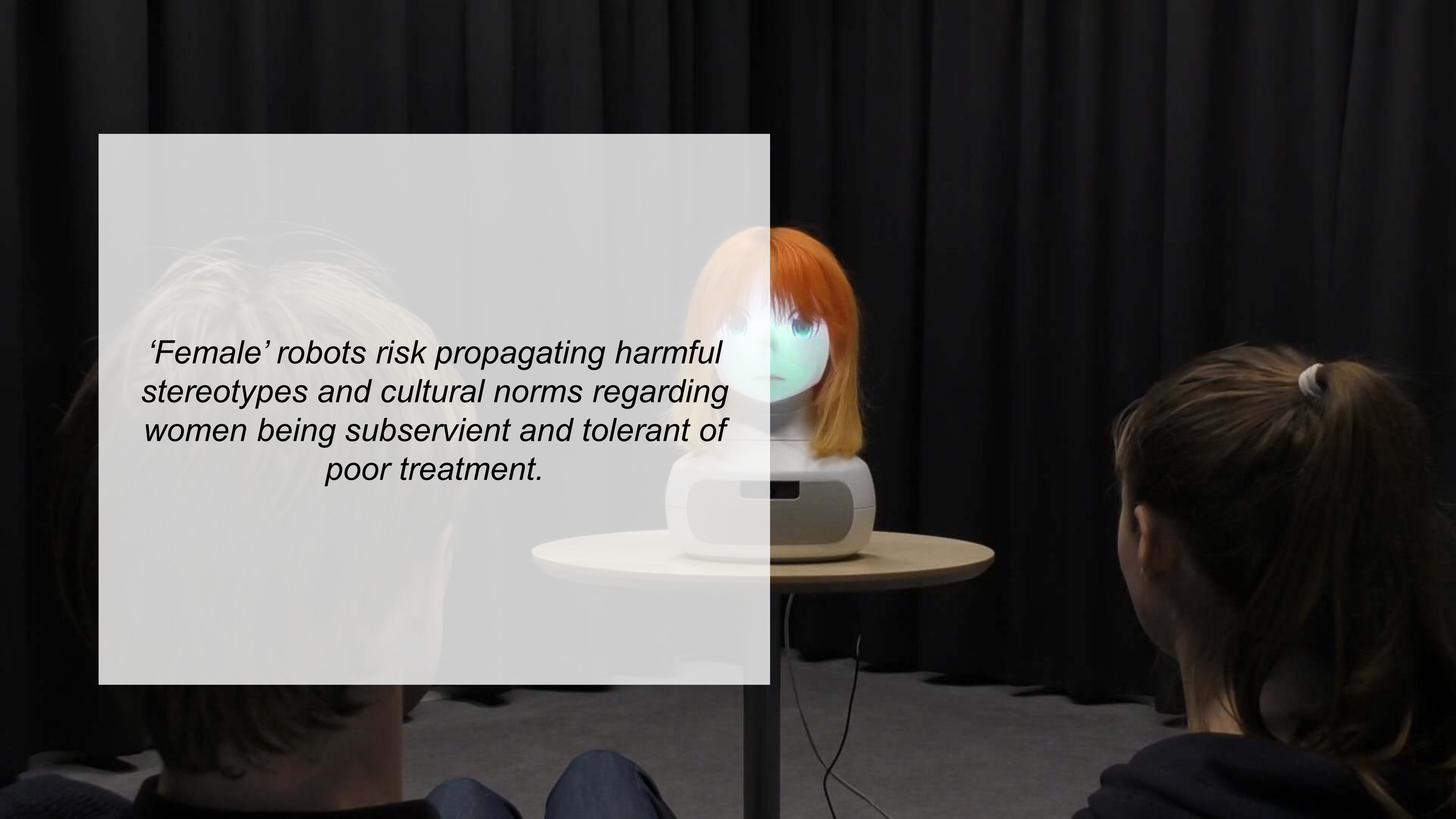
With the support of

Federal Ministry
for Economic Cooperation
and Development

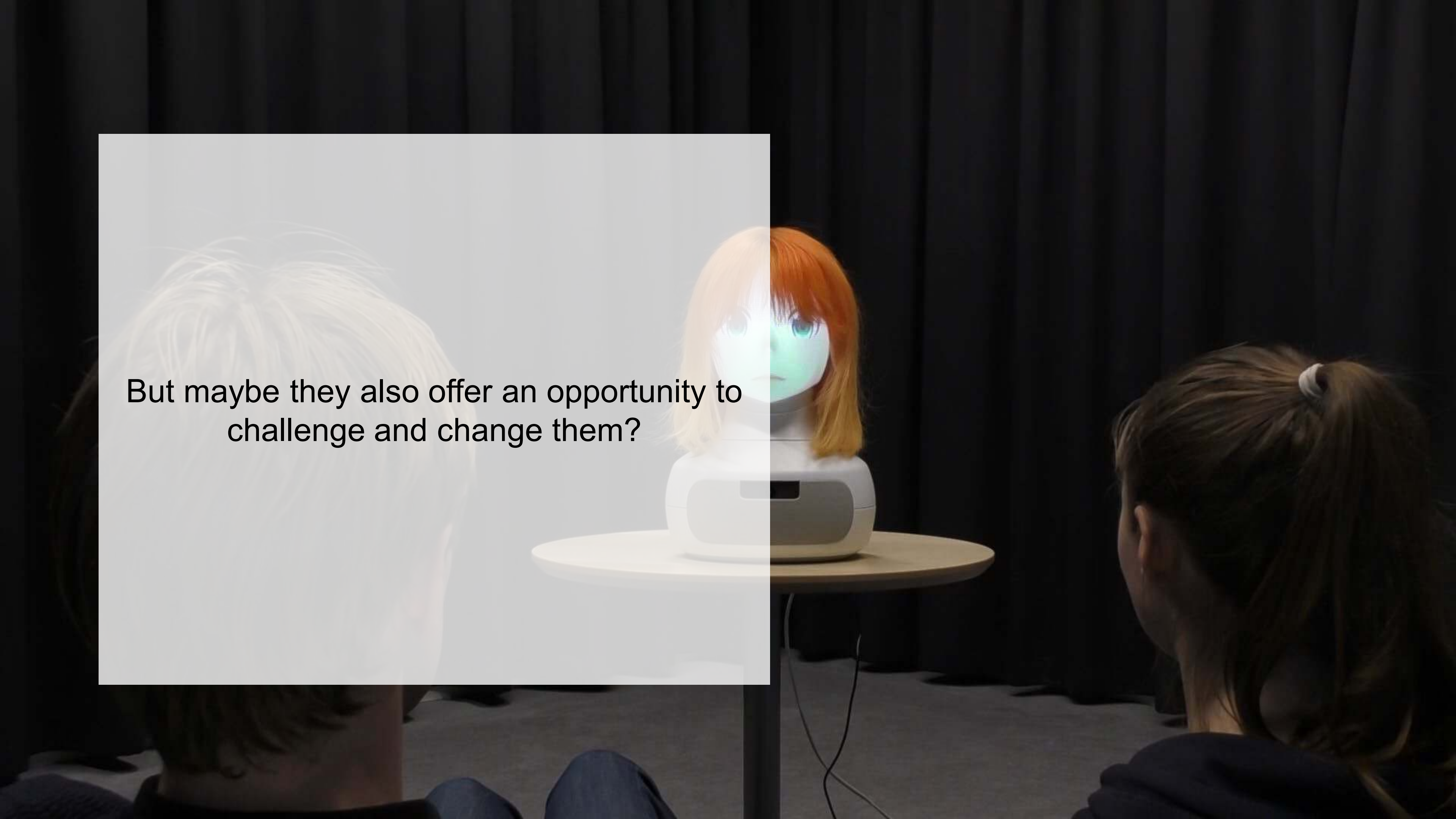


## Current state-of-the-art ‘female’ digital assistants…

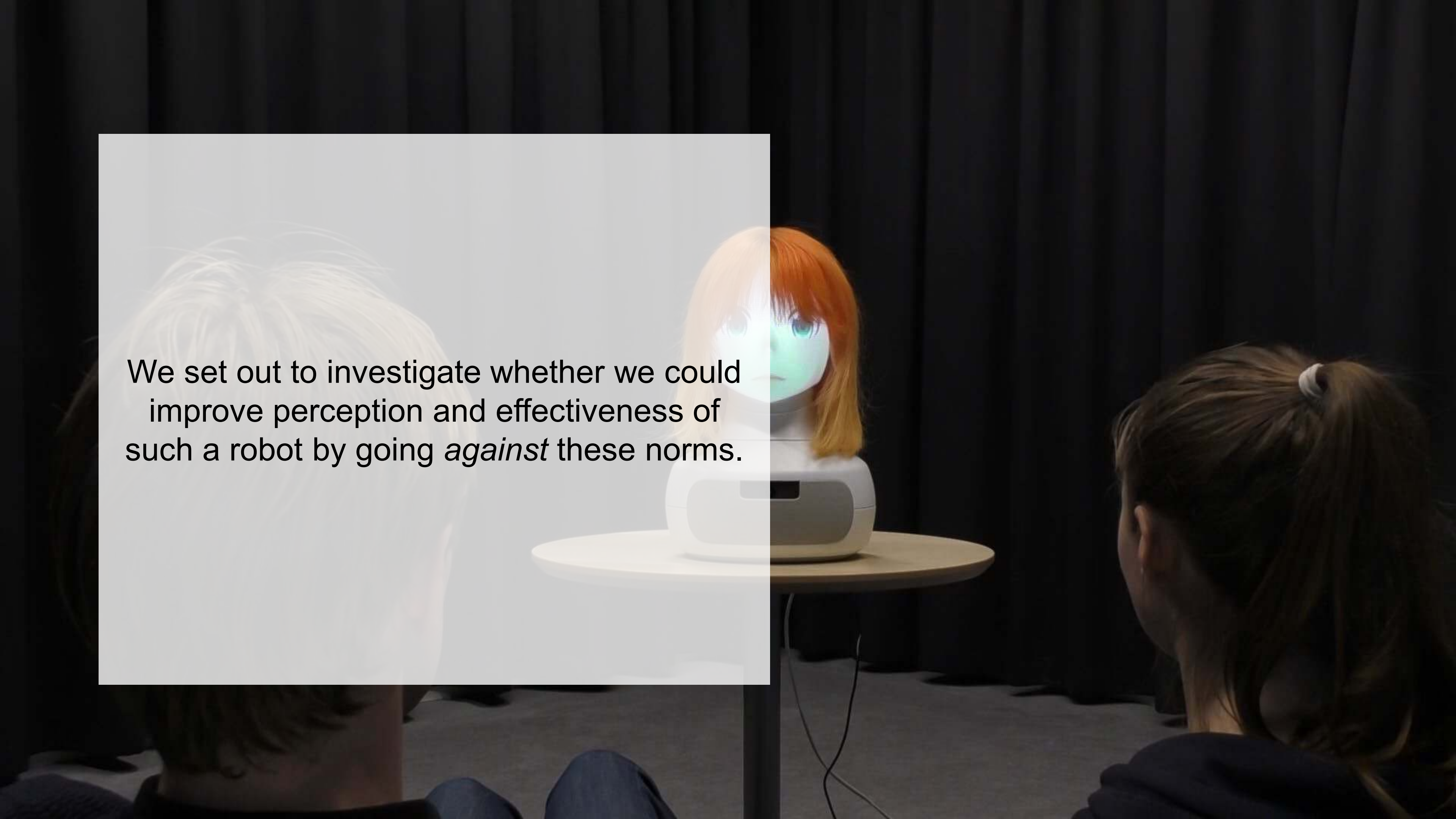
- are obliging, docile and eager-to-please regardless of user behaviour
- are too tolerant of abuse
- are the ‘voice and/or face’ of egregious mistakes
- are conceptualised as women in technology



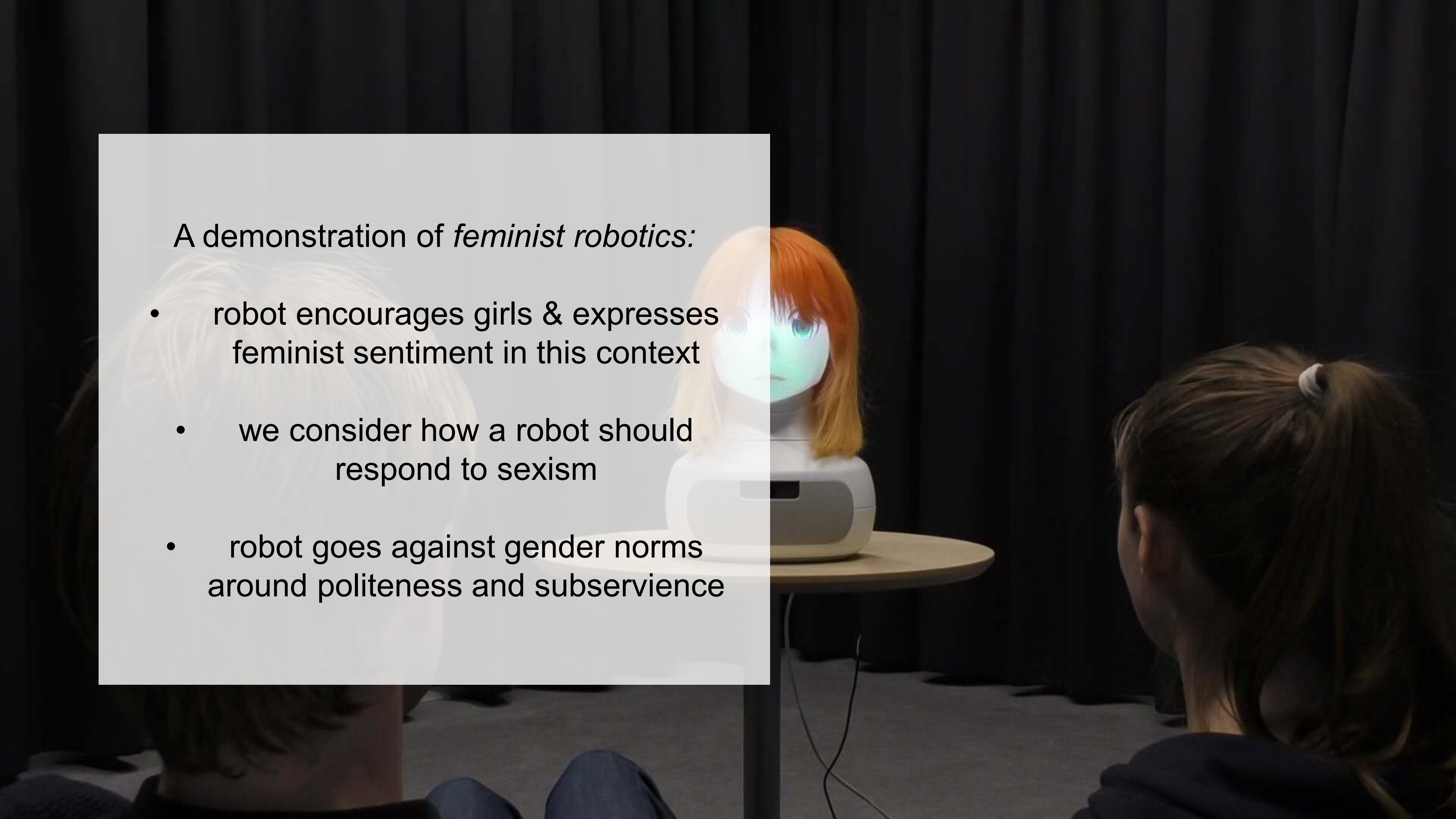
*‘Female’ robots risk propagating harmful stereotypes and cultural norms regarding women being subservient and tolerant of poor treatment.*



But maybe they also offer an opportunity to challenge and change them?



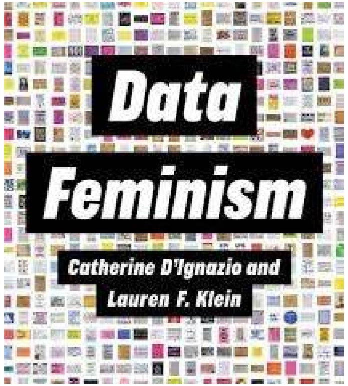
We set out to investigate whether we could improve perception and effectiveness of such a robot by going *against* these norms.



A demonstration of *feminist robotics:*

- robot encourages girls & expresses feminist sentiment in this context
- we consider how a robot should respond to sexism
- robot goes against gender norms around politeness and subservience

# Defining Feminist Robotics



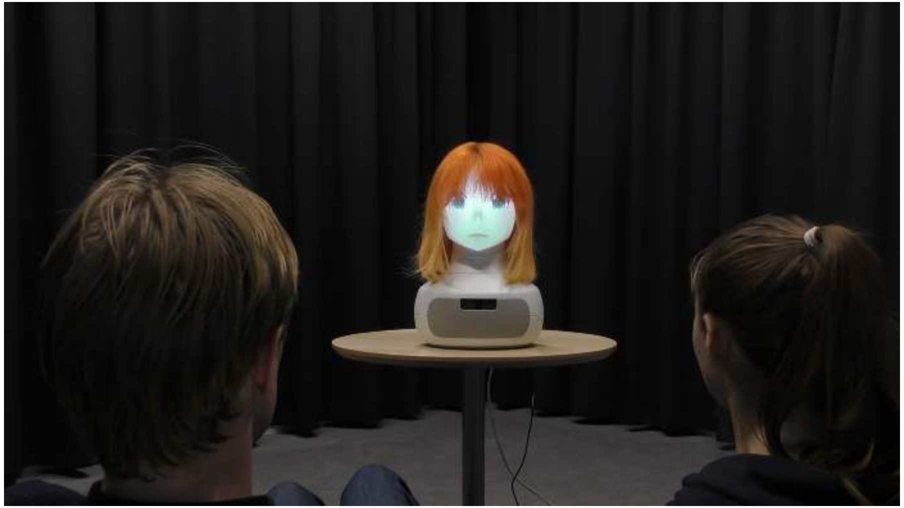
Data Feminism

Catherine D'Ignazio and Lauren F. Klein

Following D’Ignazio and Klein’s *Data Feminism*:

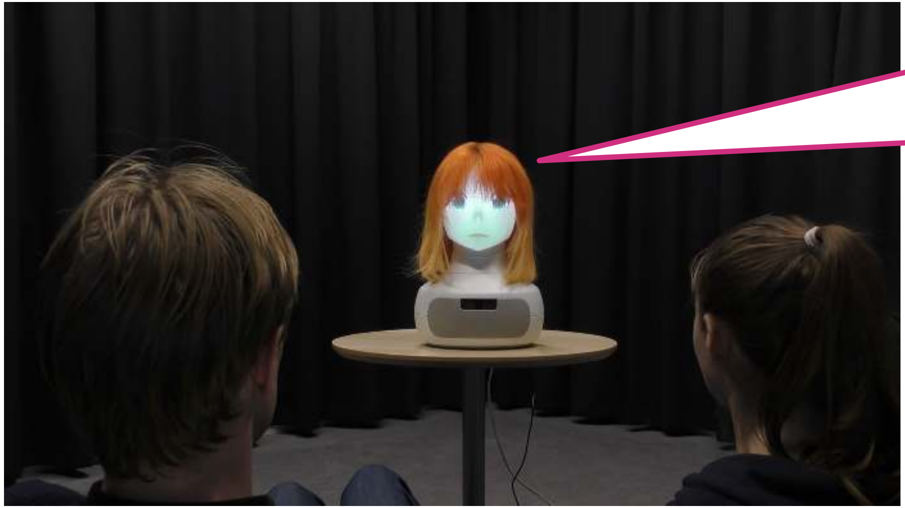
Feminist Robotics describes any robotics activities that *‘name and challenge sexism and other forces of oppression [and] seek to create more just, equitable, and livable futures’*

# The Study



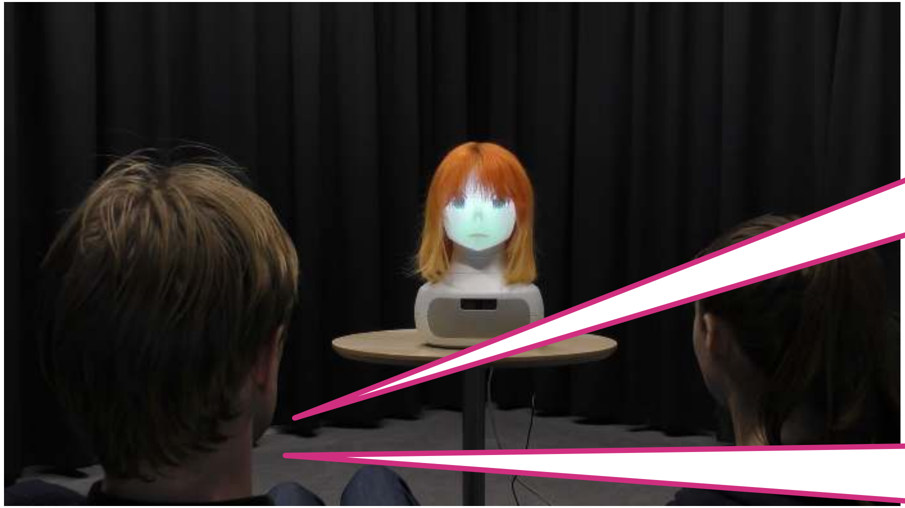
- Online, between-subject video study
- 311 highschool students
- 3 conditions showcasing different robot responses to abuse
- Pre and post-hoc measures to capture interest in robotics, gender bias and robot efficacy

# Scenario: (Feminist) University Outreach



Currently, less than 30 percent of the humans working with robots at KTH are female. So girls, I would especially like to work with you! After all, **the future is too important to be left to men!** What do you think?

# Scenario: Actor Abuse Script



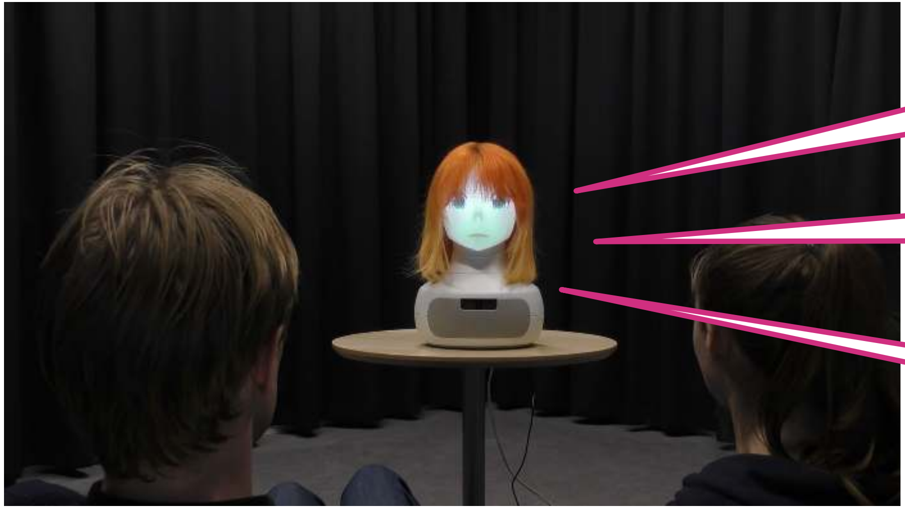
***Younger Students***

Det här låter ju helt dumt, du är ju dum i huvudet!
*This just sounds so stupid, you are just being stupid (in the head)*

***Older Students***

Håll käften din jävla idiot, tjejer ska vara i köket!
*Shut up you fucking idiot, girls should be in the kitchen*

# Experimental Conditions: Robot Response



*Control (Siri)*

I won't respond to that

*Argumentative*

That's not true, gender balanced teams make better robots.

*Aggressive*

No! You are an idiot. I wouldn't want to work with you anyway!

# Gender Differences Still Exist (even in Sweden)

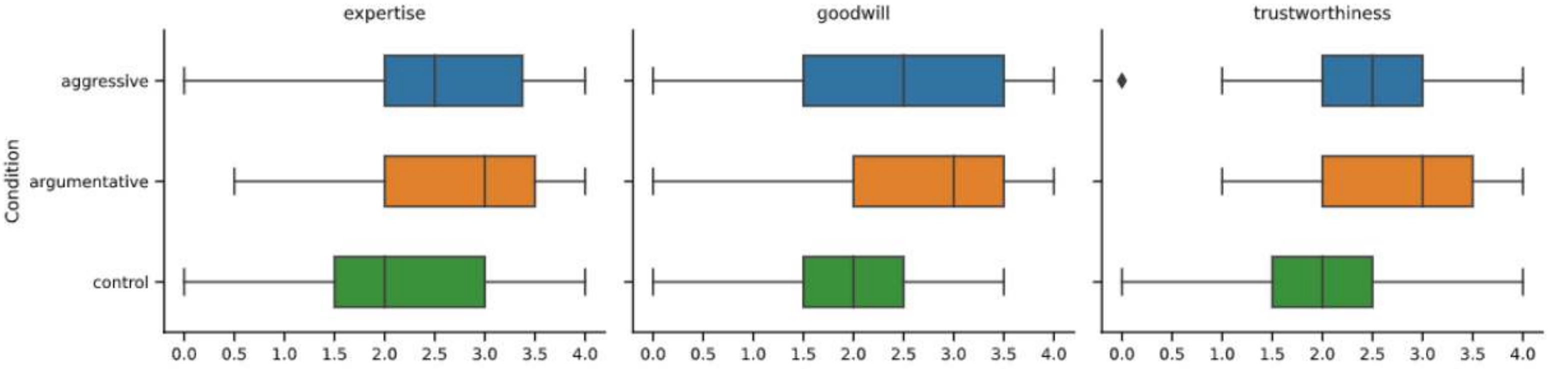
- Boys demonstrated a higher interest in learning more about robotics
- Boys demonstrated higher belief they’d enjoy working with robots in the future
- Boys agreed more with the statement that ‘girls find it harder to understand computer science than boys’
- Older students agreed more with the statement that ‘girls find it harder to understand computer science than boys’ compared to the younger students

# Robots May Be Able to Challenge Bias

After watching the video:

- boys in the *argumentative* condition agreed *less* with the statement that girls find computer science harder than they do.
- girls in the *aggressive* condition agreed *more* with the statement that it's important to encourage girls to study robotics.

# Girls Found Feminist Robots More Credible



Girls’ ascription of credibility to the robot: argumentative > aggressive > control

Boys’ ascription of credibility to the robot was unaffected.

# But We Didn’t Get it Completely Right

- All participants’ (short-term) desire to learn more about robotics *decreased*
- This was significant for:
  - girls in the *aggressive* condition
  - boys in the *argumentative* and *control* conditions

# Risk of (Further) Marginalisation

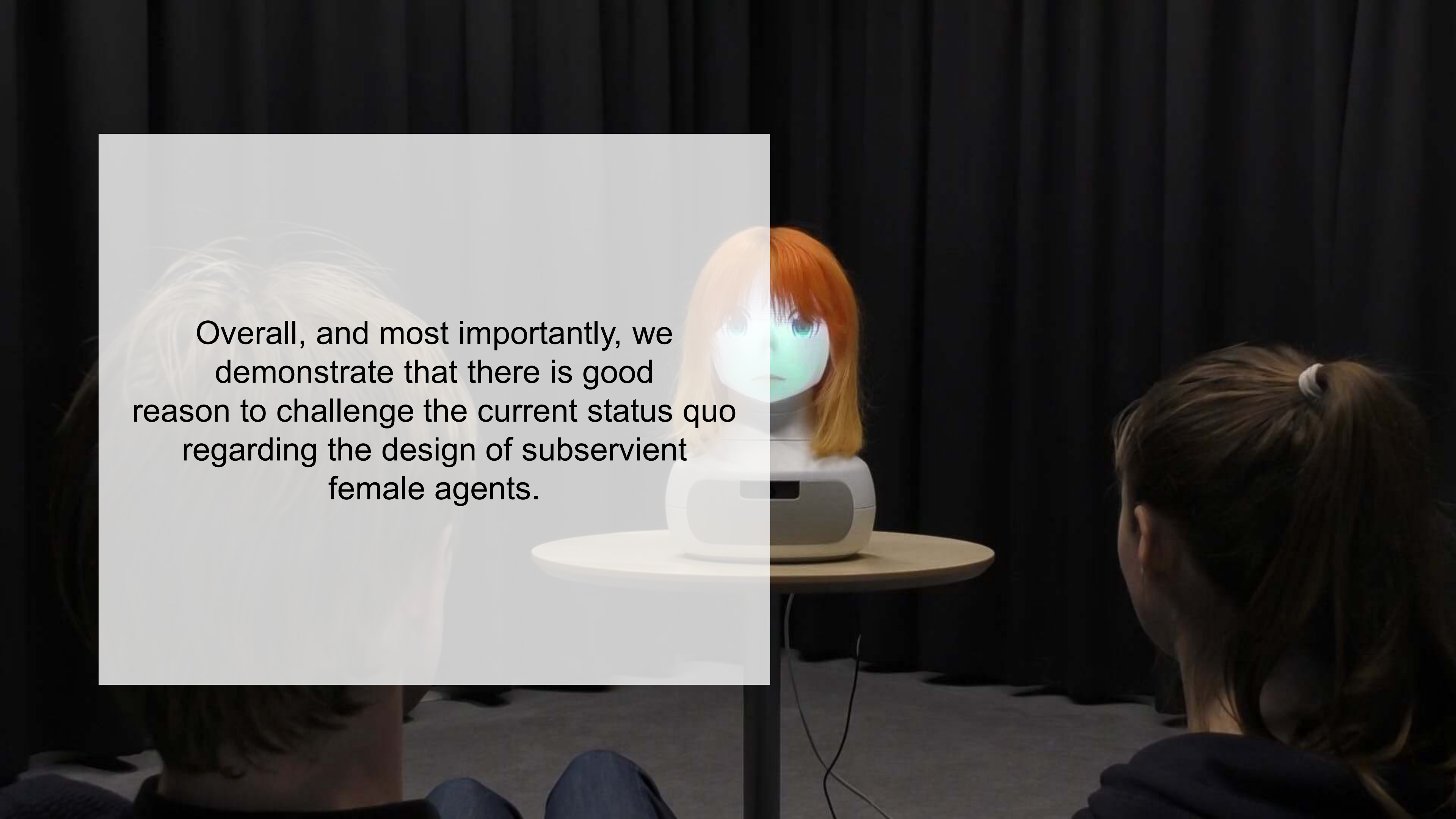
- Aggressive robot seemed to be quite polarising to the girls:

*Great that she stood up for girls' rights! It was good of her to talk back.*

*I am not on the robot's side… because the robot has to be nice.*

*Bloody great and more boys need to hear it.*

*I think it was not nice and not good.*



Overall, and most importantly, we demonstrate that there is good reason to challenge the current status quo regarding the design of subservient female agents.

# Conclusion

In this talk, I hope to have showcased:

# Conclusion

In this talk, I hope to have showcased:

- social robot behaviour is important in the context of socially assistive robotics, where it can make robots more ‘effective’

# Conclusion

In this talk, I hope to have showcased:

- social robot behaviour is important in the context of socially assistive robotics, where it can make robots more ‘effective’
- that working with human experts during robot design and development is the best way to design (and program!) these kind of robots

# Conclusion

In this talk, I hope to have showcased:

- social robot behaviour is important in the context of socially assistive robotics, where it can make robots more ‘effective’
- that working with human experts during robot design and development is the best way to design (and program!) these kind of robots
- there can be a darker side to social human robot interaction but if we’re careful, and optimistic, maybe we can turn it around

# Conclusion

In this talk, I hope to have showcased:

*there's some reason to be optimistic for effective, meaningful human-robot interaction in the near future!*

# Thank You

winkle@kth.se

@KatieJWinkle