

Value-Sensitive Design in Robotics Systems:

Ethical Challenges of Anthropomorphism

**Ethics Discussion for CS 4610/5335
Spring 2022 semester**

Vance Ricks

Associate Teaching Professor of Philosophy and Computer Science
Department of Philosophy/Religion, Northeastern University
vance@ccs.neu.edu

Topics

Today (Monday 21 March)

- + Introduction and some definitions: The tale of hitchBOT
- + Technological design and users' expectations
- + Deceptive anthropomorphic framing
- + A short introduction to VSD (value-sensitive design)

Two Mondays from today (04 April)

- + VSD continued: identifying values, stakeholders, tradeoffs
- + Applying a VSD approach to carebots

Goals

Acquire conceptual tools to help you evaluate your and others' design and uses of robotics technologies

Learn the idea of anthropomorphic framing

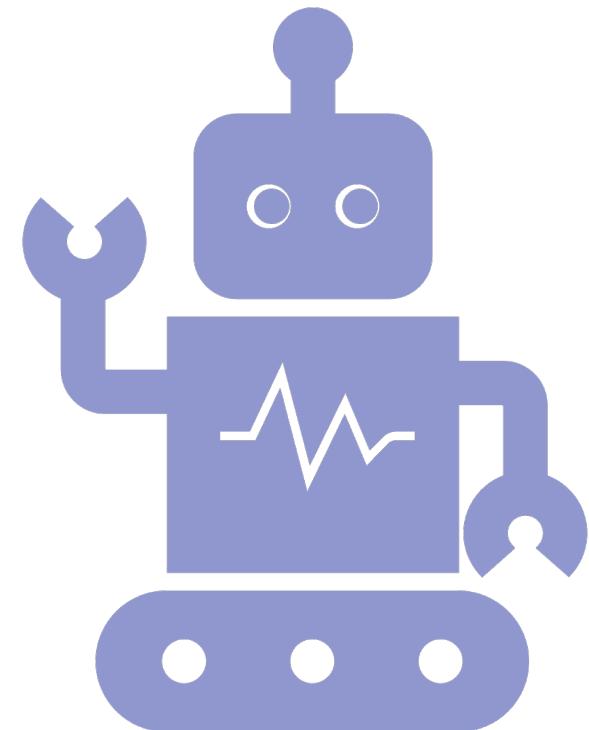
Explore some ethical implications of the design of robotics systems

Learn and **apply** value-sensitive design methods to the design of robotics systems

Evaluate the claim that technologies are inherently value-laden

Important Caveat

- + This discussion focuses on current and very-near-future robotics systems
- + No 'droids, Terminators, Ava...
- + While we can take Daniel Dennett's "intentional stance" towards these objects, we are not ascribing any sentience or consciousness to them – we are talking "as if"
- + That **still** leaves us with **many** difficult and important ethical questions!



Part 1. Introduction and some definitions: the case of hitchBOT



Source: http://www.hitchbot.me/wp-content/media/hB_MediaKit_Summer2014.pdf

What (and why) is hitchBOT?

Professors Frauke Zeller and David Harris Smith designed the robot to learn about how people interact with technology and ask the question, “Can robots trust human beings?”

HitchBot had a GPS tracker, a cellular chip and a camera. It could hold basic conversations, offering charming jests and reciting random facts. The size of a 6-year-old child, it relied on solar charging and cigarette lighters in cars for energy.

Source: Los Angeles Times, 03 August 2015

Where is hitchBOT now?

HitchBot, a talking robot built by Canadian researchers, was found maimed and battered beyond repair over the weekend in Philadelphia. It was trying to hitchhike from Massachusetts to California by relying on the kindness of human drivers. Hitchhiking took hitchBot across the length of Canada and most of Germany and the Netherlands.

But it lasted just two weeks in the U.S. and couldn't escape the East Coast.

"My trip must come to an end for now, but my love for humans will never fade," the robot said in his last Facebook message.

Source: Los Angeles Times, 03 August 2015

Four Questions To Consider

- + (When and why) Can robots trust human beings?
- + (When and why) Can human beings trust robots?
- + How should robots be designed in order to be worthy of our trust?
- + How should human beings treat robots?



2 min silent reflection, 5 minutes small-group discussion

- + **(When and why) Can robots trust human beings?**
- + **(When and why) Can human beings trust robots?**
- + **How should robots be designed in order to be worthy of our trust?**
- + **How should human beings treat robots?**



Some key definitions:

1

Robot → “a goal-oriented machine that can [semi-autonomously] sense, plan, and act” (Corke, p. 5)

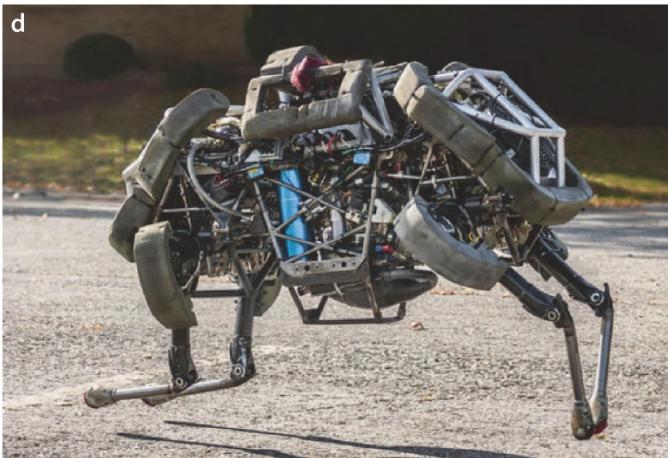
2

Social robot → a robot that does those things in the context of social interactions [with humans]

3

Anthropomorphism → the act or tendency of ascribing “human” traits (cognition, affect, perception) to nonhuman entities

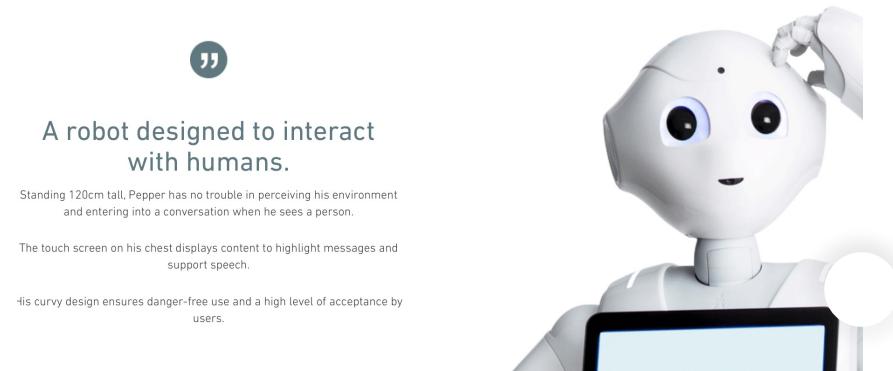
Some Varieties of Robot



Some Varieties of Social Robot

From top to bottom:

1. Paro (Paro Robots);
2. Kaspar (U of Hertfordshire);
3. Pepper (SoftBank Robotics)





What Do We Anthropomorphize?



SHIPS



CARS



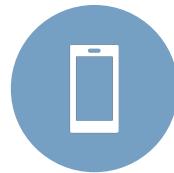
FLORA



MICROBES



WEATHER



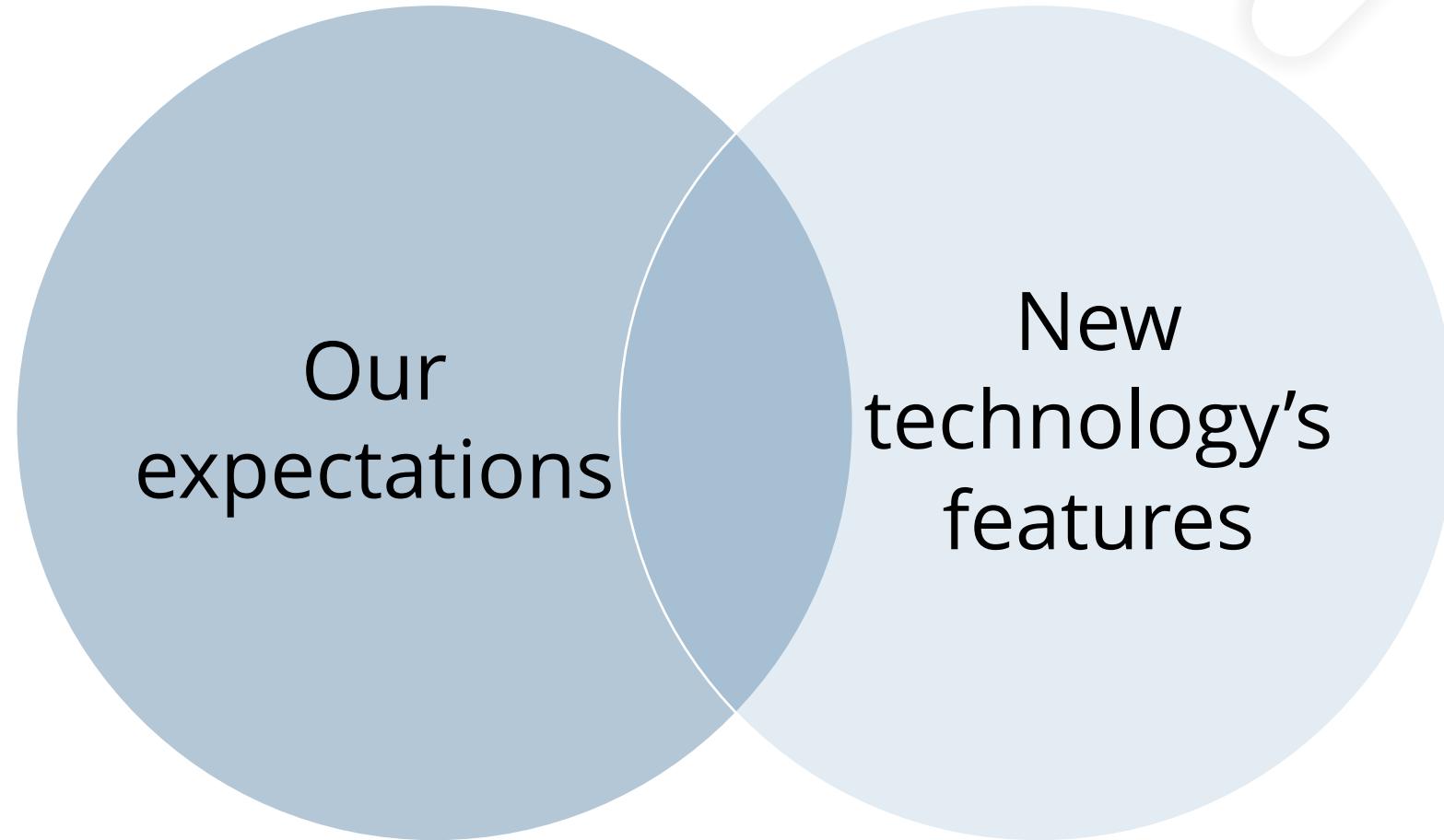
DIGITAL
DEVICES

...

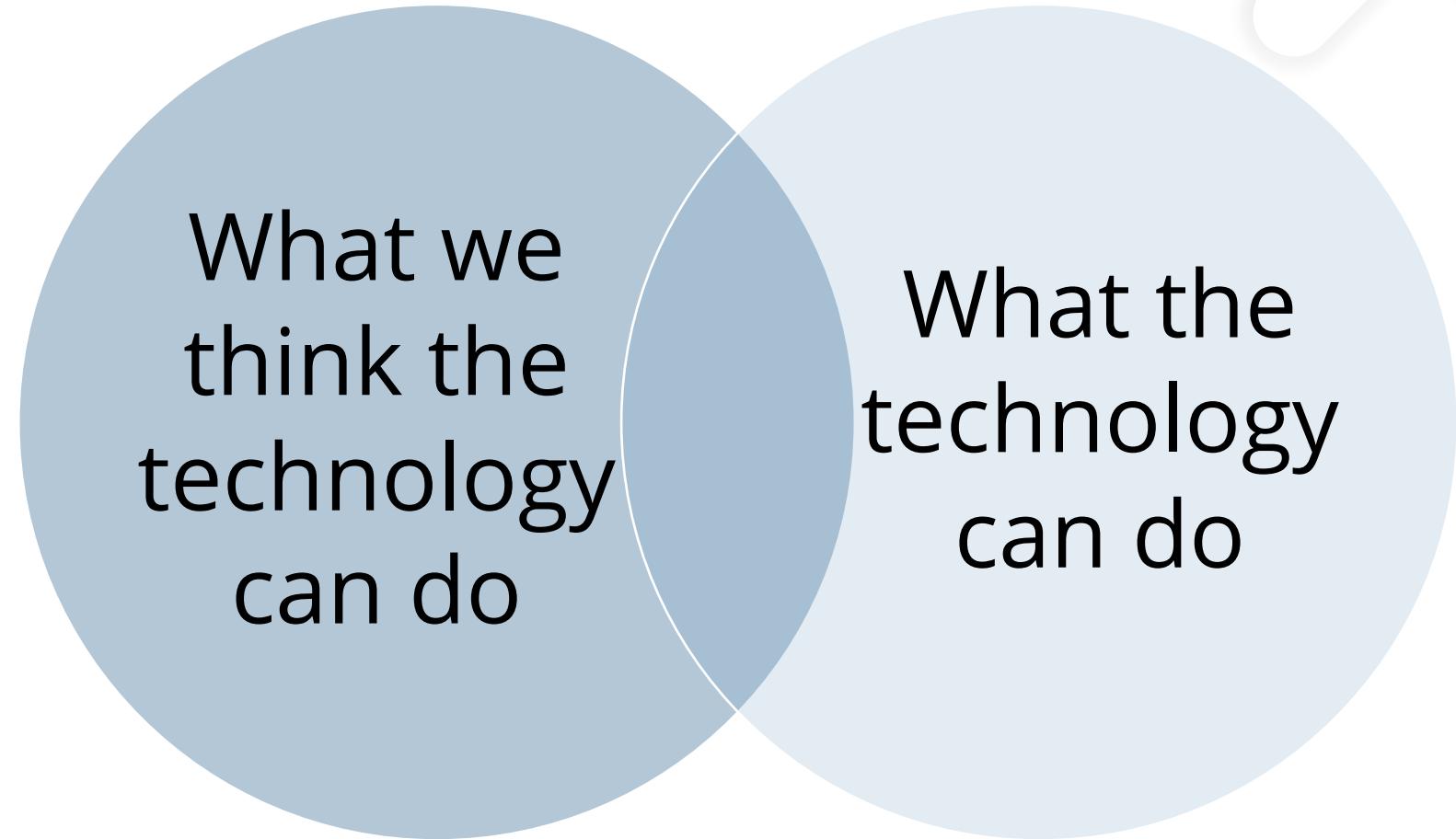
Part 2. Technological design and users' expectations



Old “boundaries” v new technological systems



Old “boundaries” v new technological systems

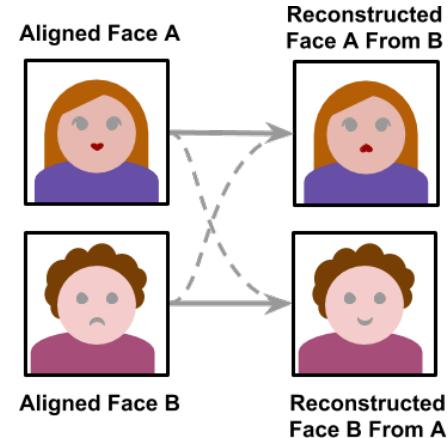


Examples of mismatches (“boundary violations”)

- + Samsung's “SmartTV”
- + phishing, catfishing, and deepfakes
- + “Fake Out Bot”



[This Photo](#) by Unknown Author is licensed under [CC BY-NC-ND](#)



[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)



[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Examples of mismatches ("boundary violations")

2 minutes individual/silent reflection
5 minutes small-group discussion

1. At least one example of a time when you underestimated a technological system's capacities:
it could do things (or be used in ways) you didn't know about
2. How did you discover that mismatch?
3. What were the consequences?



Part 3. (Deceptive) Anthropomorphic Framing

What is “framing”?



Employing a representation, description, concept, or perspective to shape perception (and actions)



Example: “Ten of you passed the midterm exam” v

“Only ten of you passed the midterm exam”



Example: Putting the least expensive foods at the back of the store



Example: Giving a tool a name

1

Using framing to encourage humans to anthropomorphize the objects they interact with – for example, by...

2

...giving an object a name

3

...giving an object a “backstory” / “biography”

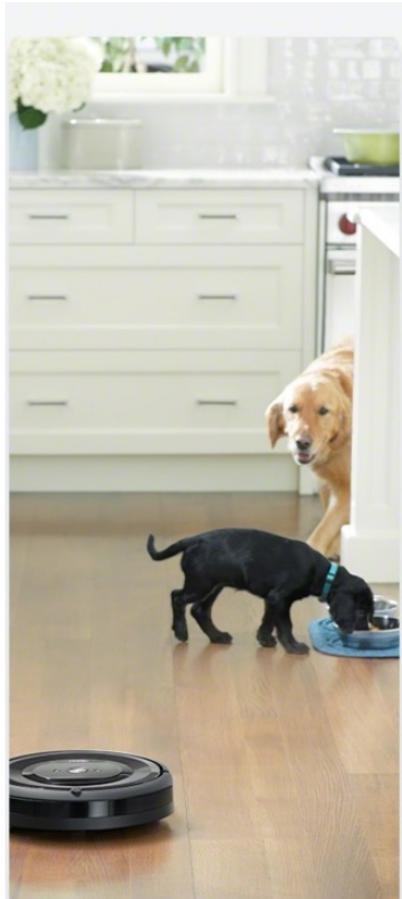
4

...giving an object specific features (“eyes”, “face”, humanoid appearance)



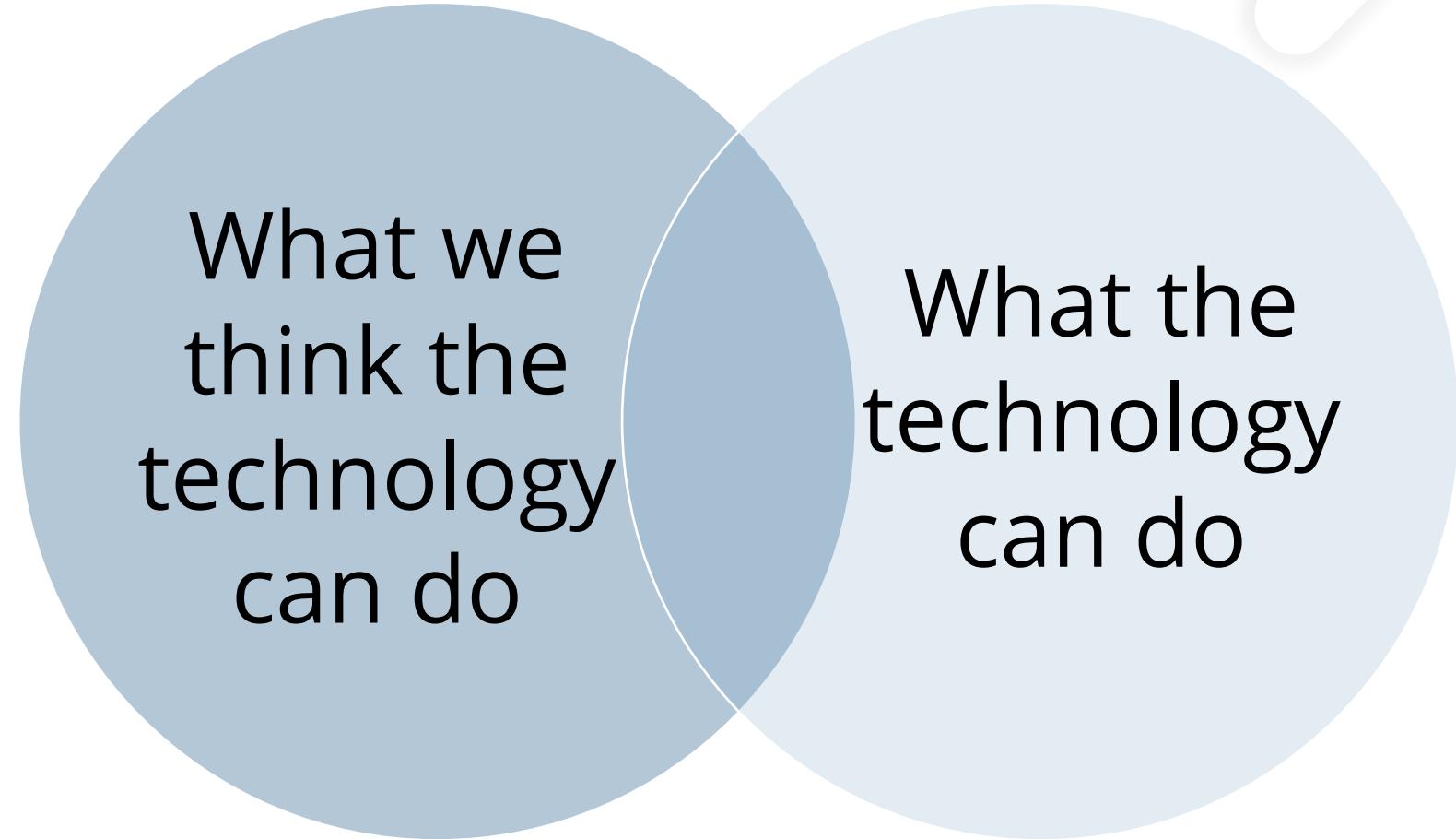
What is “anthropomorphic framing”?

Framing for “cuteness”



<https://www.irobot.com/about-irobot/smart-home>

Old “boundaries” v new technological systems



What is the risk?

Anthropomorphic framing of robots

+ boundary mismatches

= *potential for manipulation of human responses by robot designers*



Example: Roombas

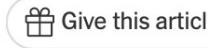
The Roomba 980 capturing & mapping its environment. Photo: iRobot²¹

Example: Boston Dynamics' "dogs"

(New York Times, 11 May 2021)

N.Y.P.D. Robot Dog's Run Is Cut Short After Fierce Backlash

The Police Department will return the device earlier than planned after critics seized on it as a dystopian example of overly aggressive policing.

 Give this article  



The Police Department used a robotic dog like this one from Boston Dynamics. The machine, which the police named Digidog, became a source of heated debate. Josh Reynolds/Associated Press

(Obligatory slide featuring a *Black Mirror* episode)

N.Y.P.D. Robot Dog's Run Is Cut Short After Fierce Backlash

The Police Department will return the device earlier than planned after critics seized on it as a dystopian example of overly aggressive policing.

 Give this article



The Police Department used a robotic dog like this one from Boston Dynamics. The machine, which the police named Digidog, became a source of heated debate. Josh Reynolds/Associated Press



CREDIT: NETFLIX

Example: gendering robots

TechCrunch, 24 Feb 2022

FOR THE RECORD

Aug. 3, 5:09 p.m.: An earlier version of this article referred to hitchBot as male. The robot was gender neutral.

Apple has developed a new Siri voice, now available in the beta versions of its iOS 15.4 software, that doesn't sound obviously male or female. The decision to introduce a gender-neutral voice is one that sees the tech giant taking yet another step away from the criticism that, historically, digital assistants have [reinforced unfair gender stereotypes](#).

Over the years, industry observers and [experts argued](#) how the creation of voice assistants with female-sounding names — like Alexa, Siri and Cortana — which also speak with female-sounding voices, implied that women should be the ones to do your bidding at any time and even [take your abuse](#). A [U.N. study](#) additionally called out the female voiced-assistants and their submissive and sometimes even flirty and coy styles.

Two Kinds of Anthropomorphism

“Honest” (Kaminski et al., 2017)

“[R]oboticists should explore using anthropomorphic features to provide better notice to users of what a robot is actually doing.”

Dishonest (Selinger & Leong, 2018)

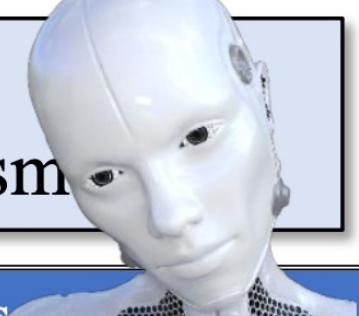
“[O]ccurs whenever the human mind’s tendency to engage in anthropomorphic reasoning and perception is abused”

A partial taxonomy of intentionally dishonest anthropomorphism

(From Selinger & Leong, 2018)

Figure 1

Taxonomy of Intentionally Dishonest Anthropomorphism



What Designers Can Abuse	Robot Examples	Possible Reasons For Deception
Human Responses to Body Parts	Installing cameras in a robot's neck or microphones in its chin.	Exploiting human expectations about embodiment in order to distort their assessment of the risks that a robot poses.
Human Responses to Body Language	Engineering a robot, that does not experience any emotions, to smile or frown.	Lulling humans into believing that a robot is friendly, hostile, sympathetic, disinterested, etc., in order to distort their assessment of the risks that it poses or nudge bonding to instill attachment or trust.
Human Responses to Physical Behavior	Engineering a robot to pretend to sleep—closed eyes, almost motionless, and curled up in the fetal position—while actively monitoring its surroundings.	Tricking humans into thinking that a robot is conveying truthful information about its behavior in order to distort their assessment of the risks that it poses.

Human Responses to Living Beings	Engineering a robot to give the physical appearance of successfully eating food that it can't digest or taste.	Manipulate humans to be invested in a robot's wellbeing, including, perhaps, nudging bonding to instill attachment or trust.
Human Responses to Size	Designing a robot to look slight and meek when it is physically powerful.	Manipulate human judgment of what a robot can do physically to distort their assessment of the risks that it poses.
Human Responses to Attractiveness	Engineering a robot to appear sexy.	Manipulate how willing humans are to please a robot and nudge bonding to instill attachment or trust.
Human Responses to Voice	Engineering a robot, that does not experience any emotion, to sound nervous or confident.	Tricking humans into thinking that a robot is conveying truthful information about its attitudes and outlooks in order to distort their assessment of the risks that it poses or nudge bonding to instill attachment or trust.

**2 min silent reflection,
5 minute small-group discussion**

(see the next slide for details)

What Designers Can Abuse	Robot Examples	Possible Reasons for Deception
Human Responses to Body Parts	Installing cameras in a robot's neck or microphones in its chin.	Exploiting human expectations about embodiment in order to distort their assessment of the risks that a robot poses.
Human Responses to Body Language	Engineering a robot, that does not experience any emotions, to smile or frown.	Lulling humans into believing that a robot is friendly, hostile, sympathetic, disinterested, etc., in order to distort their assessment of the risks that it poses or nudge bonding to instill attachment or trust.
Human Responses to Physical Behavior	Engineering a robot to pretend to sleep—closed eyes, almost motionless, and curled up in the fetal position—while actively monitoring its surroundings.	Tricking humans into thinking that a robot is conveying truthful information about its behavior in order to distort their assessment of the risks that it poses.

Do ONE of these tasks:

Either:

- Rank these from “most” to “least” concerning (give your reasons for your ranking!)

OR:

- Choose any 3 of these and give a scenario in which it could be used for intentionally deceptive purposes

Human Responses to Living Beings	Engineering a robot to give the physical appearance of successfully eating food that it can't digest or taste.	Manipulate humans to be invested in a robot's wellbeing, including, perhaps, nudging bonding to instill attachment or trust.
Human Responses to Size	Designing a robot to look slight and meek when it is physically powerful.	Manipulate human judgment of what a robot can do physically to distort their assessment of the risks that it poses.
Human Responses to Attractiveness	Engineering a robot to appear sexy.	Manipulate how willing humans are to please a robot and nudge bonding to instill attachment or trust.
Human Responses to Voice	Engineering a robot, that does not experience any emotion, to sound nervous or confident.	Tricking humans into thinking that a robot is conveying truthful information about its attitudes and outlooks in order to distort their assessment of the risks that it poses or nudge bonding to instill attachment or trust.



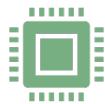
Part 4. A short introduction to Value Sensitive Design (VSD)



The case for (the need for) VSD



Technology is
the result of
human
imagination



All
technology
involves
design



All design
involves
choices
among
possible
options



All choices
reflects
values



Therefore, all
technologies
reflect and
affect human
values



Ignoring
values in the
design
process is
irresponsible

Three types of investigation in VSD

Empirical Investigation

- What do **stakeholders** say they value?
- How do stakeholders **interact with** this technology?
- What do they **know** about how the technology works?

Value Investigation

- What is the **overall goal** of the technology?
- What **values** are strengthened or undermined by this use of the technology?
- Which stakeholders are **legitimately impacted**?

Technical Investigation

- Can the tool or system be **designed** to meet value-oriented goals?
- What effect does **law, policy, and regulation have** on the design?
- Do the technical results **stay within your “red lines”**?

Value Sensitive Design (VSD) in action: the sequence

1. Who are the **stakeholders**? Identify them.
2. What **values** are at stake for those stakeholders? Identify them.
3. Where do there have to be “**tradeoffs**” between some values/interests and other values/interests?
4. Which **core ethical values** need to be given priority, or “**ethical red lines**” should not be crossed?
5. **Repeat** steps 1 – 4 as you get new information or as circumstances change.

Have a clear understanding of how the design can be **technologically successful**, not just technically successful.

WHAT OUR HISTORY
WITH ANIMALS REVEALS
ABOUT OUR FUTURE
WITH ROBOTS

THE NEW BREED



KATE DARLING

VALUE SENSITIVE DESIGN

SHAPING
TECHNOLOGY
WITH MORAL
IMAGINATION

BATYA FRIEDMAN
DAVID G. HENDRY

We must, therefore,
make careful, explicit
choices BEFORE the
fact, in our design of
our technological
systems.

This should be done not
just at the **law and
policy** level, both also
at the **design** level,
which is where value-
sensitive design – an
approach that
emphasizes
stakeholder interests
and values – attempts
to intervene.

For Next Time

Monday 04 April

- Discussion of short out-of-class assignment
- Quick review of today's materials
- Articulating values and identifying stakeholders: applying a VSD approach to carebots
- Conclusions