

Value Analysis Design: Ethics Modules in the CS Curriculum

Prepared for the Illinois Computer Science
Teaching Workshop
Tuesday 06 June 2023

Vance Ricks
Northeastern University



First, thanks

- Ron Sandler
 - John Basl
 - Matt Kopec
 - Christo Wilson
 - Kevin Mills
 - Ava Wright
 - Mark Wells
-
- The ICSSTW program committee, for having invited me!

Overview

Background

Who teaches it, and what is it?

When and how often do they teach it?

Which courses are we talking about?

Why take that approach?

What have we learned so far?

Recent Background (2018 – 2021)

Responsible Computer Science Challenge

With Great Code Comes Great Responsibility

a partnership of

 OMIDYAR NETWORK



SCHMIDT FUTURES

Craig Newmark Philanthropies

Who teaches this?



Meica Magnani

Assistant Teaching
Professor of Philosophy
and Computer Science,
Northeastern University



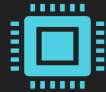
Vance Ricks

Associate Teaching
Professor of Philosophy
and Computer Science,
Northeastern University

What is it that we teach?



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
descriptively
and
prescriptively
irresponsible

What is it that we teach?

Each module = 1 or 2 class sessions (100 minutes - 200 minutes total)

1. (Introduction to the Value Sensitive Design (VSD) framework)
2. Exploration of a course-related technological system
3. Discussion of some ethical questions / problems related to that system
4. Presentation and discussion of key, relevant philosophical concepts (VAD)
5. In-class activity (application of those concepts to those questions)
6. In- (or out-of-) class activity (application of VSD framework to a case)
7. "Debriefing" and surveys (Fall 2020 - Fall 2022 semesters)



Value Sensitive Design

A framework/procedure for identifying, analyzing, and making value-based decisions in the design, development, and deployment of technology.

Three types of investigation in VSD

Empirical Investigations

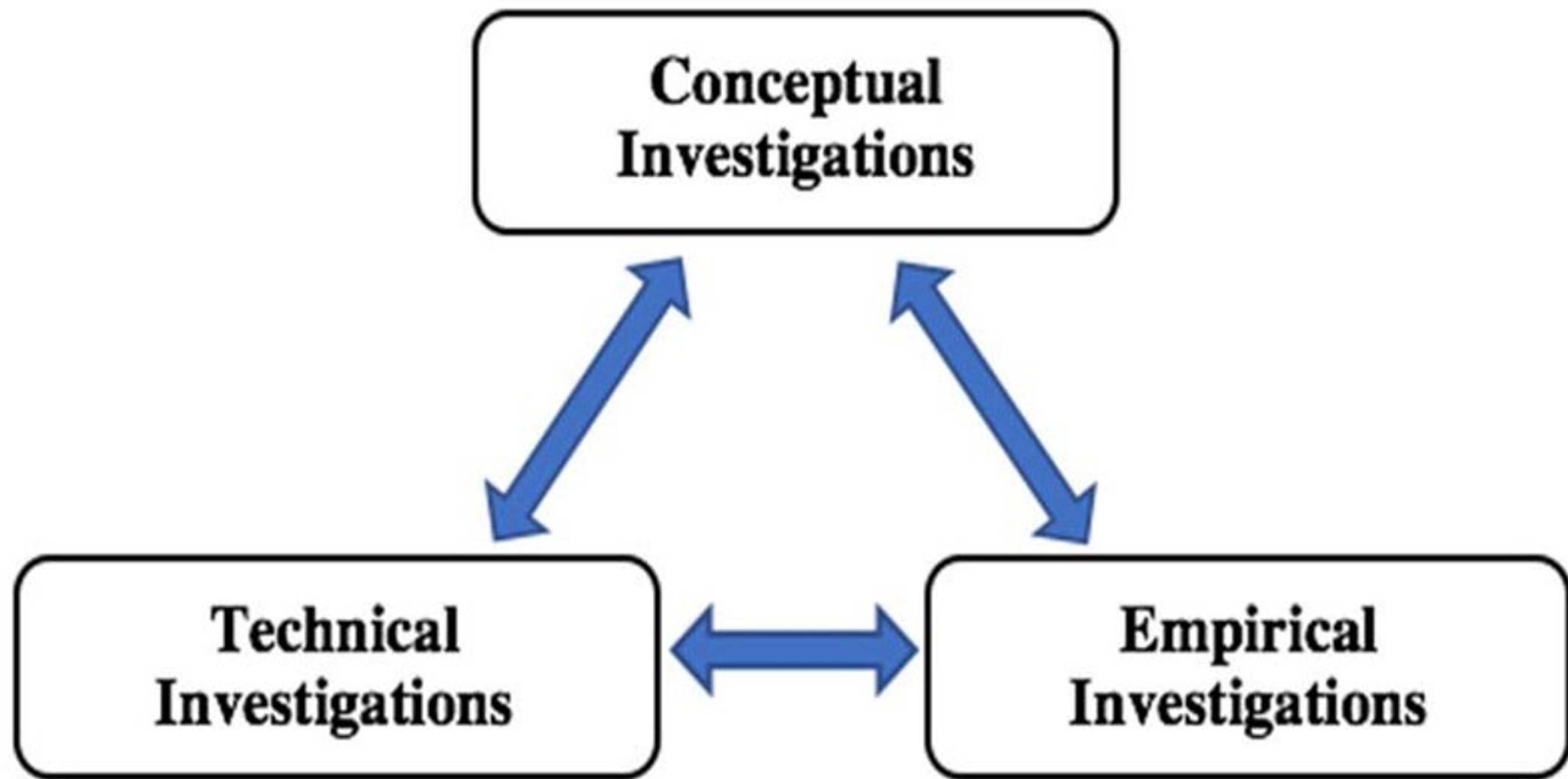
- In what **social contexts** will the technology be used?
- How do **stakeholders** interact with it?
- How do they **prioritize competing values**?
- What are the **benefits/costs** and their distributions?

Value Investigations

- What is the **overall goal** of the technology?
- What **values** are at stake?
- How do we **refine our understanding** of those values in this context?
- Which **legitimate stakeholders** are affected?
- What value-based criteria will be used to **gauge project success**?

Technical Investigations

- What frameworks and tools enable designers to meet value-based goals?
- What effects do **law, policy, and regulation** have on the design?
- Do the technical results **stay within ethical boundaries**?



Framing the Technical Work



Framing the Technical Work

- 
1. Clarify the project values from the engineer/designer stance
 2. Situate the work within a social context

Values/Conceptual Investigations

What/who are the stakeholders?

What are their interests/values?

Are their interests/values aligned?

How **should** we **understand** these values?



Empirical Investigations

What sort of empirical information about the world (psychological, social, cultural, biological, etc.) is needed in this context?

What practices or norms are **already** present?

What social and environmental conditions surround materials and production?

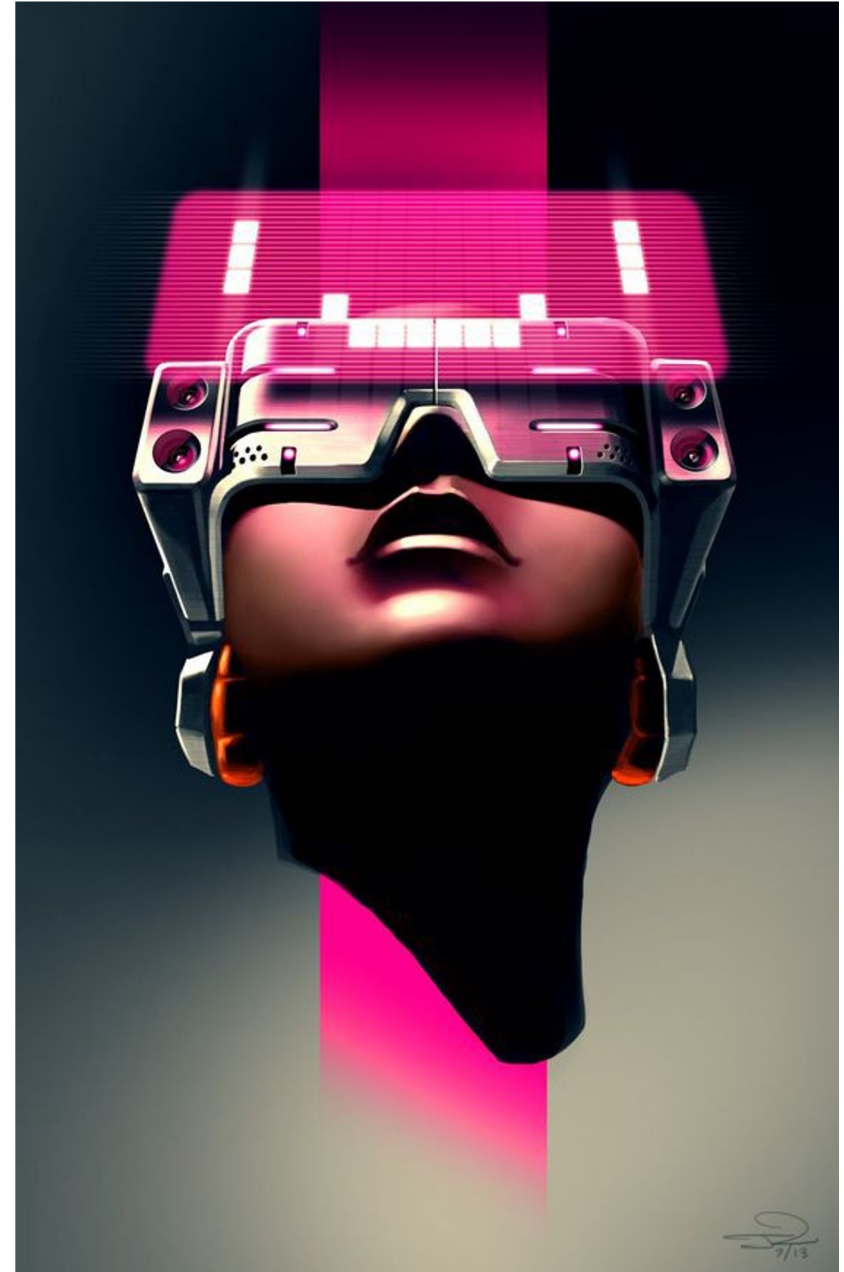


Technical Investigations

What sort of trade-offs might there be between technical optimization and implicated values?

How does the technology interact with current norms and practices?

Should new regulations be introduced?



**Value
Sensitive
Design (VSD)
in action:
the sequence**

1. Who are the **stakeholders**? Identify them.

2. What **ethical values** are at stake for those stakeholders?
Identify them.

3. Where do there have to be “**tradeoffs**” between some ethical values (or interests) and other ethical values (or 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.

When and how often does it occur?

Fall and Spring semesters (from Fall 2019 – present)

Earlier = better; ideally within the first month or first half of term

One (preferably two) class visits, sequential or within a week

Which courses are we talking about?

CS, Cybersecurity, and DS courses (Fall 2019 – present):

- Artificial Intelligence
- Natural Language Processing
- Web Development
- Robotics Science and Systems
- Algorithmic Auditing
- Foundations of Cybersecurity
- Machine Learning/Data Mining 1 and 2

UD/UX courses et al. (tbd)

Why take that approach?

Modularity

Embedded in existing CS courses = easier to fit within “crammed” CS curriculum

Tailored to address course-specific ethical issues = more engaging to students*

Taught by competent and motivated instructors = more engaging to students*

Multiple modules offered across CS curriculum = repeated exposure to applied ethical concepts and skills throughout their course of study*

VSD-inspired

Emphasizes inescapability of values in design/development (not just implementation) of technologies

Sees technologies as parts of sociotechnical and sociocultural systems, not as “neutral tools”

Highlights potential sociopolitical implications of design choices

Encourages multidisciplinary, multiperspectival approach to weighing moral implications of those choices

Example 1 (NLP course)

EXAMPLE 1: NATURAL LANGUAGE PROCESSING

Session 1

Activity: word analogies generated by word embeddings

Small group discussion: nature of stereotypes, what makes them ethically problematic

Values analysis: allocative vs representational harm

Discussion: implicit bias, stereotype threat, positive stereotype



EXAMPLE 1: NATURAL LANGUAGE PROCESSING

Session 2

Students introduced to the VSD framework — value, empirical, and technical investigations (including stakeholder analysis).

Guided VAD analysis of hypothetical NLP application (ranking candidates based on letters of recommendation), with the aim of anticipating and mitigating bias and harms.



Example 2 (Robotics Science & Systems Course)

Example 2: Robotics Science and Systems

Session 1

- **Activity:** How should robots be designed in order to be worthy of our trust?
- **Small group discussion:** How should we treat robots?
- **Values analysis:** care; “honest” v “dishonest” anthropomorphism
- **Discussion:** care; legitimate and illegitimate forms of deception and manipulation

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

Example 2: Robotics Science and Systems

Session 2

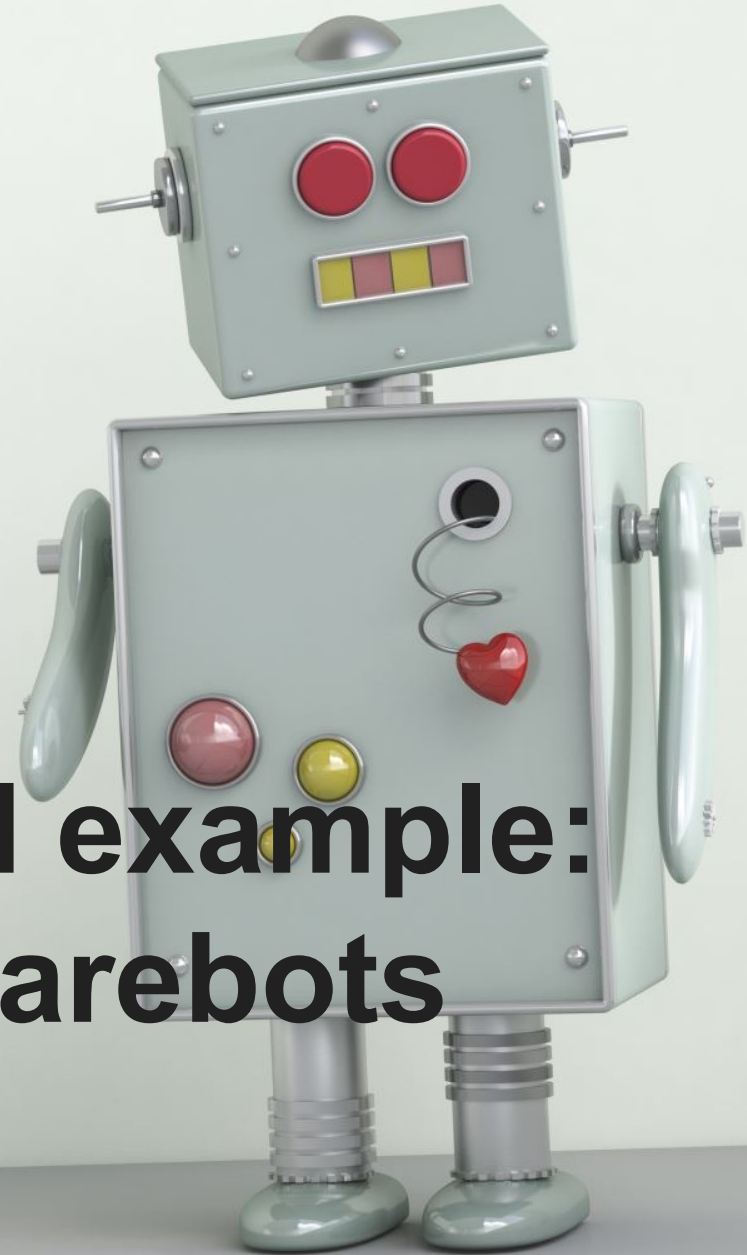
Students introduced to the VSD framework

Guided VAD analysis of a hypothetical robotic system application (designing a carebot), with the goals of

- (a) avoiding dishonest anthropomorphism and
- (b) not undermining human responsibilities for caretaking



**A more detailed example:
the case of carebots**



What is “care”?

Caring about (affective)

- An attitude
- An internal state
- Feeling of concern (for well-being, safety, comfort, etc.)

Caring for (functional)

- An activity or set of activities
- Not necessarily indicative of any particular internal state
- Doing things to safeguard, protect, assist, etc.

What are “carebots”?

1. Embodied social robots designed for use in home, hospital, or other settings
2. Embodied social robots whose function is to assist in, support, and/or provide **care** for sick, disabled, young, elderly, or otherwise vulnerable people (Vallor 2016)
3. “(Social) robots intended for inclusion in the daily **care activities** of persons – like lifting, feeding, or bathing – used by a care-giver in the care of another [person] or used by the care-receiver directly” (van Wynsberghe 2013)

**Examples of
carebots**

MySpoon

Paro

ElliQ

Riba

Rudy

Considering Carebots

1. Assistive, enabling, or replacing?
2. Deceptive?
3. Invasive and/or insecure?
4. Manipulative?



Considering Carebots

What values or interests might make carebots a step in a good direction?

- Ageing populations → fewer strong younger people to care for them
- Round-the-clock care that doesn't get bored or tired
- Greater information storage → more reliable care
- \$
- Shortage of skilled human caregivers
- Less neglect, abuse, incompetence
- Effects on caregiver-caretaker relations
- Greater safety and reliability when lifting, etc.
- Making caretaking less gendered and less fraught

What values or interests might make carebots a step in a bad direction?

- Loss of valuable human interactions
- Effect on caregiver-caretaker relations
- Risk of dishonest anthropomorphism
- Manipulative technologies
- Deceptive technologies
- Expectations mismatch
- So much sensitive information collected!

Designing a carebot

–
applying a VSD
approach



Group up – your task is to **design a carebot**



In your group, decide which **context(s) of use** you have in mind for your carebot



Decide which **care practices** you want your carebot to perform or contribute to

Three types of investigation in VSD

Empirical Investigation

- How might carebots **assist or replace** human caretakers?
- How do people *providing* care **define** “care”?
- How might people’s **expectations** of *receiving* care change if carebots become common?

Value Investigation

- Is care an ethical value, or an **umbrella term** for a range of ethical values?
- Does expressing care require things that **only humans** can (best) do or provide?
- When might the use of carebots demonstrate a **lack** of care?

Technical Investigation

- Can we make “honestly anthropomorphic” carebots?
- How should carebots be designed to preserve patients’ sensitive personal information?
- How can carebots pick up nonverbal cues?

VAD approach – a checklist

**How will it exemplify
some aspect of
either affective care
or functional care?
(Be specific!)**

**What should it
be able to do
or to help
someone else
do?**

**What must it NOT do
or help someone
else do?**

**How will it avoid
crossing your ethical
“red lines”?**

An outlook, not a (rote) algorithm

Fundamentally, VSD is an outlook and a process

VSD is not an algorithm

There is no design recipe for VSD

There is no way to *#include vsd.h* or *import VSD*

Committing to VSD means being thoughtful and responsive

No single right answer to complex ethical and moral questions...

But there are **lots of wrong answers**

Engaging with values in the design process offers creative opportunities for:

- Technical innovation
- Improving the human condition (*doing good and saving the world*)

What Have We Learned So Far? Reflections and Future Directions

REFLECTIONS: ON MODULES

STRENGTHS

- Signals that ethics is part of CS
- Connects ethics and technical content
- Demonstrates concrete application of philosophical tools
- Brings discussion into the CS classroom

WEAKNESSES

- Risks overemphasizing designers' role
- Ethical* values must be made explicit
- Ethicist may still seem like an outsider
- Hard to assess outcomes

REFLECTIONS: VALUES ANALYSIS DESIGN

STRENGTHS

- Provides a framework to guide ethical analysis and intervention
- Shows the interdisciplinary nature of designing responsible technology
- Highlights ways in which values can get in at every step of the way

WEAKNESSES

- Can be quite a bit of (seeming at first) jargon to throw at students
- Not clear students will remember this methodology after a couple sessions
- A lot to cram into a session or two
- Presumes clear investigative tools and normative value framework

FUTURE DIRECTIONS:

ADDRESSING WEAKNESSES

- role playing with detailed case studies

- pre-recorded video with introduction to VSD

- further integration

FURTHER INTEGRATION

- link content/concepts in modules to content/concepts in Technology and Human Values (required course for CS students)

- speaker series (people from industry, research, policy, community-based groups)

VAD in a CS ethics “ecosystem”

Ethics-enriched
CS intro course
sequence

VAD modules in
mid- / upper-
level CS
courses

Technology and
Human Values
course

Experiential
(incl. co-op)
programs

To learn more:

1. VSD Co-op Wiki: https://vsdcoop.ischool.uw.edu/index.php/VSD_Coop
2. (in-need-of-update) Northeastern VSD page: <https://vsd.ccs.neu.edu/>
2. Kopec, Magnani, Ricks, et al., “The Effectiveness of Embedded Values Analysis Modules in Computer Science Education: An Empirical Study”, forthcoming in *Big Data and Society*

Questions and Discussion