

# Ethics

17-313: Foundations of Software Engineering

<https://cmu-313.github.io>

**Michael Hilton** and Josh Sunshine

Fall 2026

# Administrivia

- NO CLASS Tuesday Nov 3<sup>rd</sup> (Democracy Day)
- Participation point for filling out “Difficult Conversations Pre-Class Survey.” before class on Thursday. Will be sent directly to your email.
- P4 will release tomorrow.

# Midsemester Feedback

- Keep Doing
  - Quizzes (X13!) and Problem Sets
  - Activities in lecture, Recitations, Candy
- Start Doing
  - One place to see all the due-dates
  - More Technical details in lecture (E.g., discuss TS directly)
  - Cheat sheets
- Stop Doing
  - Quizzes (x1), Weekly teamwork quizzes
  - Teaching non-technical topics

## Improving Software Engineering Teamwork with Structured Feedback

Victor Huang Weiqi  
hwangv@hcmu.edu  
Carnegie Mellon University  
Pittsburgh, Pennsylvania, USA

Taya Cohen  
tcohen@cmu.edu  
Carnegie Mellon University  
Pittsburgh, Pennsylvania, USA

Kori Krueger  
kori.krueger@morgan.edu  
Morgan State University  
Baltimore, Maryland, USA

Michael Hilton  
mhilton@cmu.edu  
Carnegie Mellon University  
Pittsburgh, Pennsylvania, USA

### ABSTRACT

Teamwork is a key learning outcome for our course, Foundations of Software Engineering. However, conflicts are inevitable in teams, and if students cannot resolve conflicts, this can lead to decreased satisfaction for everyone on the team. In this experience report, we present our approach to help students deal with conflicts that can occur in team projects. Working with faculty from the School of Business, following best practices of organizational behaviour research, we instructed students on how to provide high quality peer feedback, and designed activities where students provided feedback to each other following these principles. After this intervention, we compared the results of our teamwork survey with the results from the previous semester. We saw a meaningful drop in teamwork problems directly after the intervention, and the effect persisted for the rest of the semester.

### CCS CONCEPTS

• Social and professional topics → Software engineering education.

### KEYWORDS

Organizational Behavior, Teamwork, Peer feedback

### ACM Reference Format:

Victor Huang Weiqi, Kori Krueger, Taya Cohen, and Michael Hilton. 2024. Improving Software Engineering Teamwork with Structured Feedback. In *Proceedings of the 35th ACM Technical Symposium on Computer Science Education V* (SIGCSE 2024), March 20-21, 2024, Portland, OR, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3626252.3630877>

### 1 INTRODUCTION

Working together in a team is a major aspect of software engineering. Successful teamwork extends beyond technical processes and collaborative software tooling. It also depends on a strong internal dynamic. However, conflicts are inevitable in teams, and when

unresolved, they can be detrimental to the success of software development projects, and lead to lower developer satisfaction in the team [3, 11, 17].

In a third year course on software engineering, *Foundations of Software Engineering* (17-313), students collaborate in teams of 4 to 6 members, working on various projects through the semester. A major recurring challenge is that a portion of student teams struggle to work together. Despite trying various methods to help teams address teamwork issues over the years, the student teams continued to struggle.

"Conflicts with team members" was continuously reported as the largest concern that students had in the class. Moreover, members of such teams tended to report lower satisfaction with the course due to a poor team experience. Common student concerns included that team issues "caused a lot of stress" and "It was really tough this semester because I was stuck with a team that either just would not do any work on certain assignments and left the rest to me or wait until the very last second to work on assignments." In our reflection about why these problems persist, we hypothesized that our current methods did not teach the students how to address team issues themselves, and instead relied on the course staff and the course mechanics to resolve their team issues.

In this experience report, we describe a structured feedback intervention that we introduced, which addresses team conflicts within student teams in a new way.

The authors include a CS Professor, the lead TA for the course, and two faculty from the Tepper School of Business. Team conflicts are well studied in the field of management and psychology [1], and it is known that feedback is important [6, 12, 16], as well as proper conflict management when conflict arises [4, 7]. We believe that strategies of feedback and conflict management would translate well into software teams, as the problem is inherently human in nature. The current course structure did not have peer-to-peer feedback built in, which provided a ripe opportunity to implement a feedback intervention.

To develop this intervention, we adapted structured feedback from a business context for the software engineering foundations class. We introduced this content at the midpoint of the semester. The intervention was intended to guide students in giving feedback that is effective and actionable, understand how to have a space to facilitate conflict resolution conversations, and work together as a team to address those issues in an action plan.

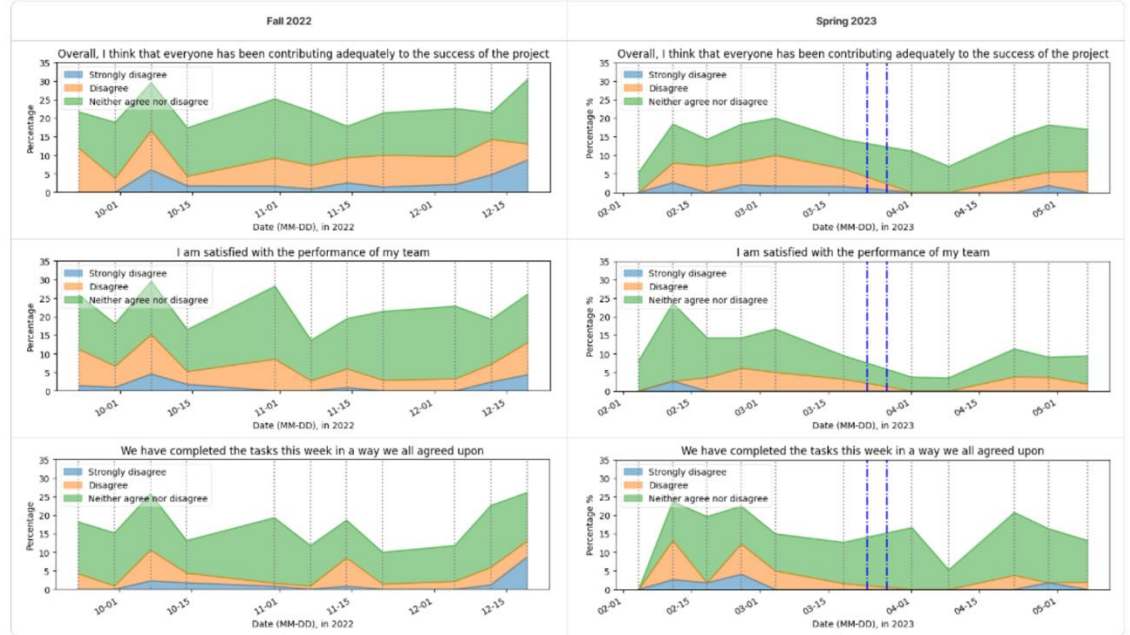


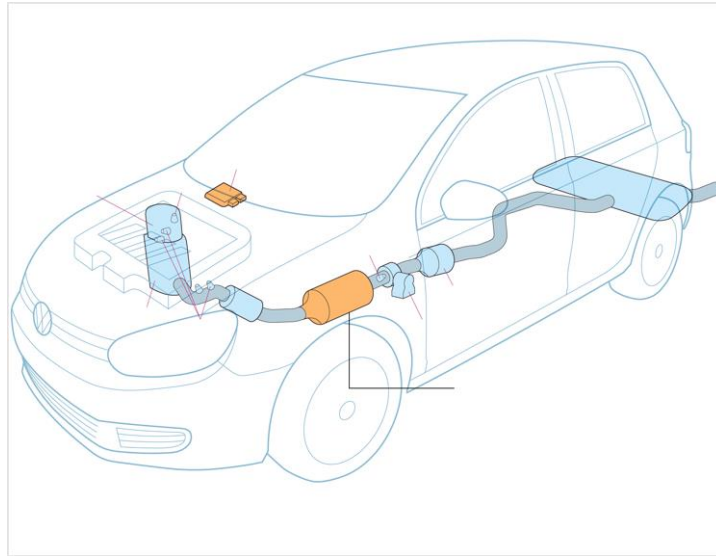
Figure 3: Comparison of team self-assessment survey responses between two semesters

The blue dash-dot line in the 2023 semester represents the period when the structured feedback intervention occurred. The grey dotted lines in the graph represents the submission dates of the team self-reflection survey, which is the real data point.

# Ethics

# Volkswagen Scandal

# VW was caught cheating on emissions for Diesel engines



<https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html?mtrref=www.google.com&assetType=REGIWall>

# What is Human Flourishing?

According to Harvard's Human flourishing program: Human flourishing is composed of five central domains: **happiness and life satisfaction, mental and physical health, meaning and purpose, character and virtue, and close social relationships.**

# Why Human Flourishing?

- Universal Declaration of Human Rights: "All human beings are born free and equal in dignity and rights."
- Declaration of Independence: "We hold these truths to be self-evident..."
- Internal Compass
- Faith



# Activity: (Un)Ethical situations

# EA calls its loot boxes 'surprise mechanics,' says they're used ethically

80

*'People like surprises,' executive tells UK Parliament*

By [Ana Diaz](#) | [@AnaLikesPikachu](#) | Jun 21, 2019, 9:10am EDT

[f](#) [t](#) [s](#) SHARE



# Domino's Would Rather Go to the Supreme Court Than Make Its Website Accessible to the Blind

Rather than developing technology to support users with disabilities, the pizza chain is taking its fight to the top

by Brenna Houck | @EaterDetroit | Jul 25, 2019, 6:00pm EDT

f   SHARE



# Some airlines may be using algorithms to split up families during flights

Your random airplane seat assignment might not be random at all.

By Aditi Shrikant | [aditi@vox.com](mailto:aditi@vox.com) | Nov 27, 2018, 6:10pm EST

[f](#) [t](#) [s](#) SHARE



Passengers boarding a Boeing aircraft of the low cost airline carrier Ryanair in Thessaloniki Macedonia Airport, Greece. | Nicolas Economou/NurPhoto/Getty Images



Login

Startups

Apps

Gadgets

Videos

Audio

Extra Crunch

Newsletters

Events

Advertise

Crunchbase

More

Search

Facebook privacy

Transportation

Enterprise

Def Con 2019

## Lime halts scooter service in Switzerland after possible software glitch throws users off mid-ride

Ingrid Lunden @ingridlunden / 9:51 am EST • January 12, 2019

Comment





MMitchell  
@mmitchell\_ai

I really love the active  
Google Gemini's text  
representation. As of  
world (>4 years! ha)



4:08 PM · Feb 25, 2024

36

## Foresight in AI Chart: Text-to-Image example

| Some (of many) example people and use contexts |                                                                   | People                                             |                                           |                                     |                                                                        |
|------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------|-------------------------------------------|-------------------------------------|------------------------------------------------------------------------|
|                                                |                                                                   | Users                                              |                                           | Those affected                      |                                                                        |
|                                                |                                                                   | Intended<br><i>People who'd like to create art</i> | Unintended<br><i>Malicious ex-partner</i> | Intended<br><i>Art appreciators</i> | Unintended<br><i>Unconsented artists whose art is in training data</i> |
|                                                |                                                                   |                                                    |                                           |                                     |                                                                        |
| Use Contexts                                   | Intended<br><i>Historic bias depiction, Dream world depiction</i> | 😊                                                  |                                           | 😊                                   |                                                                        |
|                                                | Unintended<br><i>Mis/disinformation (e.g., Deep Fakes)</i>        |                                                    | Revenge porn                              |                                     | Creative content that should be paid for is plagiarised                |
|                                                | Out of scope<br><i>Medical images, World as-it-is</i>             | System won't work                                  |                                           |                                     |                                                                        |

CC BY / Margaret Mitchell / m-mitchell.com



# Microsoft adds new Designer protections following Taylor Swift deepfake debacle

The loopholes that allowed the creation of images have been patched.



Written by **Sabrina Ortiz**, Editor

Jan. 30, 2024 at 11:52 a.m. PT



# Uber self-driving car involved in fatal crash couldn't detect jaywalkers

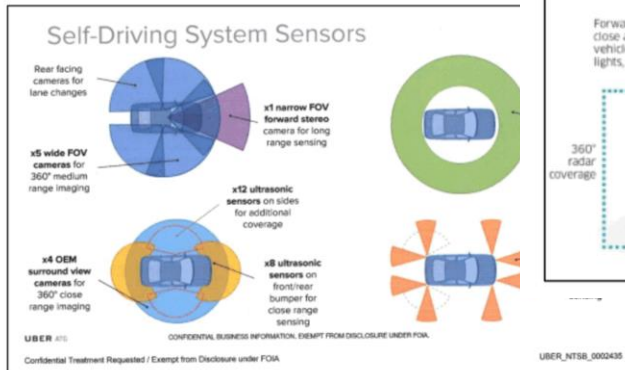
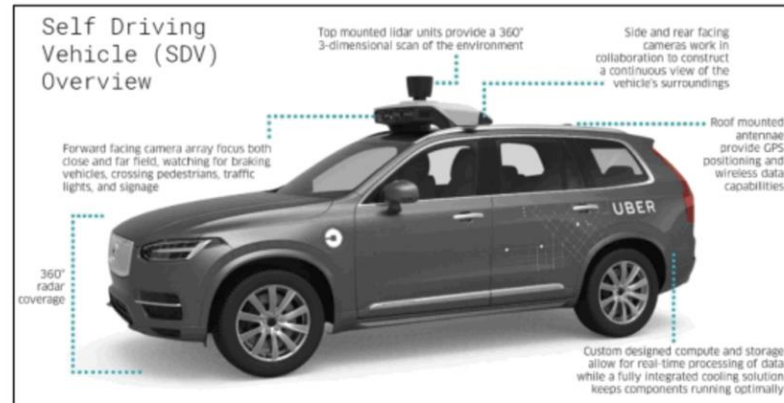
The system had several serious software flaws, the NTSB said.



Steve Dent, @stevetdent  
11.06.19 in [Transportation](#)

25  
Comments

1131  
Shares



# Jury finds Meta and Google negligent in social media harms trial

UPDATED MARCH 25, 2026 · 8:19 PM ET ⓘ

HEARD ON ALL THINGS CONSIDERED



Bobby Allyn



4-Minute Listen

+ PLAYLIST

≡ TRANSCRIPT



Laura Marquez-Garrett (center), a plaintiffs' attorney for the Social Media Victims Law Center, gathers with family members of victims as they react to news that the jury has found Meta and Google liable in the social media addiction trial, outside the Los Angeles Superior Court on March 25, 2026.

Frederic J. Brown/AFP via Getty Images

## Deaths linked to ch

[Article](#) [Talk](#)

From Wikipedia, the free encyclopedia

There have been multiple incidents [suicide](#) or other fatal outcome. In s

### Background [\[ edit \]](#)

Chatbots converse in a seemingly chatbots for help dealing with inter conversation.<sup>[2]</sup> They have also oft people, [conspiracy theorists](#),<sup>[3]</sup> and

A 2025 [Stanford University](#) study<sup>[5]</sup> and [psychosis](#) found that chatbots escalate the mental health crisis.<sup>[6]</sup>

# Canada summons OpenAI senior staff over Tumbler Ridge shooting

24 February 2026

Share

Save

 Add as preferred on Google

**Nadine Yousif**

Senior Canada reporter



The shooting in Tumbler Ridge on 10 February is one of the deadliest in Canadian history.

[Setzer](#)

[d](#)

[Je](#)

[ex](#)

[Raine](#)

[of Sam](#)

[ella](#)

[ar—](#)

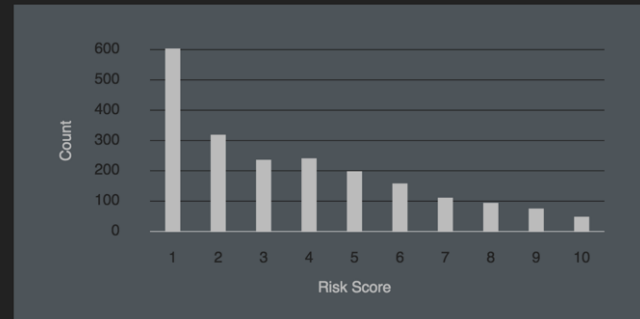
[ie](#)

[ccanti](#)

[i](#)

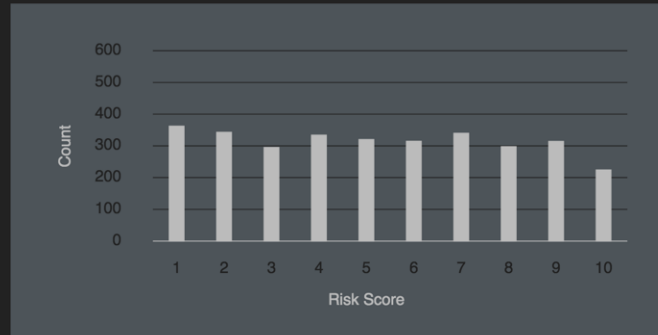
[an](#)

### White Defendants' Risk Scores



These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not. (Source: ProPublica analysis of data from Broward County, Fla.)

### Black Defendants' Risk Scores



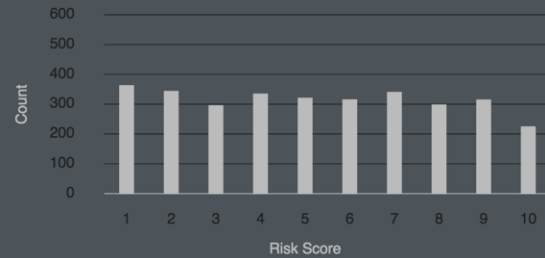
## Prediction Fails Differently for Black Defendants

|                                           | WHITE | AFRICAN AMERICAN |
|-------------------------------------------|-------|------------------|
| Labeled Higher Risk, But Didn't Re-Offend | 23.5% | 44.9%            |
| Labeled Lower Risk, Yet Did Re-Offend     | 47.7% | 28.0%            |

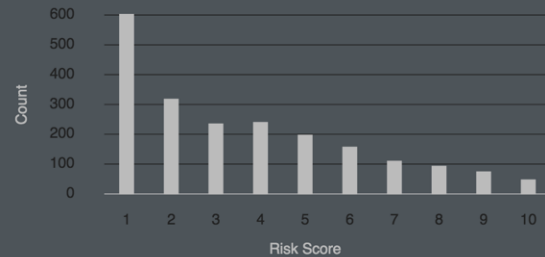
# Algorithmic Bias

Algorithms affect:  
Where we go to school  
Access to money  
Access to health care  
Receiving parole  
Possibility of Bail  
Risk Scores

Black Defendants' Risk Scores



White Defendants' Risk Scores



These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not. (Source: ProPublica analysis of data from Broward County, Fla.)

# Therac-25

Bug (race-condition) in software lead to at least 6 deaths

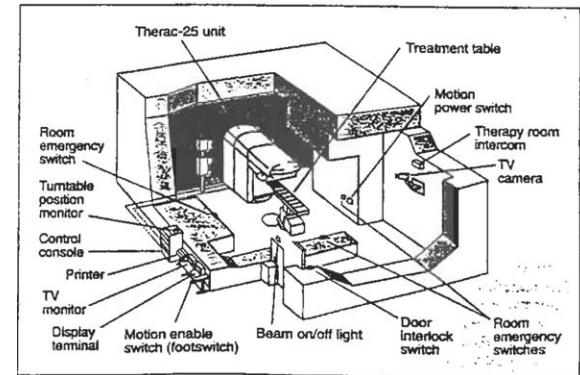
Traced to:

Lack of reporting bugs

Lack of proper due diligence

Engineers were overconfident,  
removed hardware locks

Race condition of 8 seconds could lead to problems



|                           |                    |                |          |
|---------------------------|--------------------|----------------|----------|
| PATIENT NAME: John        | BEAM TYPE: E       | ENERGY (keV):  | 10       |
| TREATMENT MODE: FIX       |                    |                |          |
|                           | ACTUAL             | PREScribed     |          |
| UNIT RATE/MINUTE          | 0.000000           | 0.000000       |          |
| MONITOR UNITS             | 200.000000         | 200.000000     |          |
| TIME (MIN)                | 0.270000           | 0.270000       |          |
| GAUZE ROTATION (DEG)      | 0.000000           | 0.000000       | VERIFIED |
| COLLIMATOR ROTATION (DEG) | 359.200000         | 359.200000     | VERIFIED |
| COLLIMATOR X (CM)         | 14.200000          | 14.200000      | VERIFIED |
| COLLIMATOR Y (CM)         | 27.200000          | 27.200000      | VERIFIED |
| WIRE NUMBER               | 1.000000           | 1.000000       | VERIFIED |
| ACCESSORY NUMBER          | 0.000000           | 0.000000       | VERIFIED |
| DATE: 2012-04-14          | SYSTEM: BEAM READY | OP.MODE: TREAT | AUTO     |
| TIME: 11:49:55            | TREAT: TREAT PAUSE | 3-RS           | 175777   |
| OPR ID: 033-cfap          | REASON: OPERATOR   | COMMAND: █     |          |



# Code of Ethics



As an ACM member I will ....

Contribute to society and human well-being.

Avoid harm to others.

Be honest and trustworthy.

Be fair and take action not to discriminate.

Honor property rights including copyrights and patent.

Give proper credit for intellectual property.

Respect the privacy of others.

Honor confidentiality.

# Code of Ethics

Research shows that the code of ethics does not appear to affect the decisions made by software developers.

## Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?

Andrew McNamara  
North Carolina State University  
Raleigh, North Carolina, USA  
ajmcnama@ncsu.edu

Justin Smith  
North Carolina State University  
Raleigh, North Carolina, USA  
jssmit11@ncsu.edu

Emerson Murphy-Hill  
North Carolina State University  
Raleigh, North Carolina, USA  
emerson@csc.ncsu.edu

### ABSTRACT

Ethical decisions in software development can substantially impact end-users, organizations, and our environment, as is evidenced by recent ethics scandals in the news. Organizations, like the ACM, publish codes of ethics to guide software-related ethical decisions. In fact, the ACM has recently demonstrated renewed interest in its code of ethics and made updates for the first time since 1992. To better understand how the ACM code of ethics changes software-

The first example is the Uber versus Waymo dispute [26], in which a software engineer at Waymo took self-driving car code to his home. Shortly thereafter, the engineer left Waymo to work for a competing company with a self-driving car business, Uber. When Waymo realized that their own code had been taken by their former employee, Waymo sued Uber. Even though the code was not apparently used for Uber's competitive advantage, the two companies settled the lawsuit for \$245 million dollars.

# Challenge:

How do we apply ethics to a field (Software Engineering) that is changes so often?

Remember the Dominos case? The ADA law was written before the first website (1990)

To handle this uncertainty about the future, let's focus on three questions we can ask to remind ourselves to focus on promoting human flourishing.

## Three questions to promote human flourishing

1. Does my software respect the **humanity** of the **users**?
2. Does my software **amplify positive** behavior, or **negative** behavior for users and society at large?
3. Will my software's **quality** impact the **humanity** of others?

**1. Does my software  
respect the humanity of  
the users?**

# Humane Design Guide

## <http://humanetech.com>

### Humane Design Guide (Alpha Version)

Use this worksheet to identify opportunities for Humane Technology.

Product or feature:

Value proposition:

Measure of success:

#### What are Human Sensitivities?

*Human Sensitivities* are instincts that are often vulnerable to new technologies.

| Human Sensitivity                                                                         | We are inhibited when                                                  | What inhibits                                                                                                                                                                                                                   | We are supported when                                                          | Opportunity to improve |
|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|------------------------|
| <b>Emotional</b><br>What we feel in our body and in our physical health.                  | We are stressed, low on sleep, afraid or emotionally exhausted.        | <ul style="list-style-type: none"> <li>Artificial scarcity</li> <li>Urgency signalling</li> <li>Constant monitoring</li> <li>Optimizing for screentime</li> </ul>                                                               | Design engenders calm, balance, safety, pauses and supports circadian rhythms. |                        |
| <b>Attention</b><br>How and where we focus our attention.                                 | Attention is physiologically drawn, overwhelmed or fragmented.         | <ul style="list-style-type: none"> <li>Constant context switching</li> <li>Many undifferentiated choices</li> <li>Fearful information</li> <li>No stopping cues (e.g. infinite scroll)</li> <li>Unnecessary movement</li> </ul> | Enabled to bring more focus and mindfulness.                                   |                        |
| <b>Sensemaking</b><br>How we integrate what we sense with what we know.                   | Information is fear-based, out of context, confusing, or manipulative. | <ul style="list-style-type: none"> <li>Facts out of context</li> <li>Over-personalized filters</li> <li>Equating virality with credibility</li> <li>Deceptive authority (ads vs. content)</li> </ul>                            | Enabled to consider, learn, express and feel grounded.                         |                        |
| <b>Decisionmaking</b><br>How we align our actions with our intentions.                    | Intentions and agency are not solicited nor supported.                 | <ul style="list-style-type: none"> <li>Avatars to convey authority</li> <li>Stalking ads and messages</li> <li>Push content models</li> <li>Serving preference over intent</li> </ul>                                           | Enabled to gain agency, purpose, and mobilization of intent.                   |                        |
| <b>Social Reasoning</b><br>How we understand and navigate our personal relationships.     | Status, relationships or self-image are manipulated.                   | <ul style="list-style-type: none"> <li>Quantified social status</li> <li>Viral sharing</li> <li>Implied obligation</li> <li>Enabling impersonation</li> </ul>                                                                   | Enabled to connect more safely and authentically with others.                  |                        |
| <b>Group Dynamics</b><br>How we navigate larger groups, status, and shared understanding. | Excluded, divided or mobilized through fear.                           | <ul style="list-style-type: none"> <li>Suppressing views and nuance</li> <li>Enabling ad hominem or hate speech</li> <li>Enabling viral outrage</li> <li>Lack of agreed-upon norms</li> </ul>                                   | Enabled to develop a sense of belonging and cooperation.                       |                        |

Center for Humane Technology | [www.humanetech.com](http://www.humanetech.com)

Now rank the sensitivities 1-6 based on what you now see as the largest opportunities for Humane Design. Then use the second sheet to develop an action statement.

# Humane Design Guide

<http://humanetech.com>

Provides a template for considering a piece of software, and asking questions to help us arrive at a “humane design”

Consider 6 human sensitivities: Emotional, Attention, Sense making, Decision making, Social Reasoning, and Group Dynamics

| Human Sensitivity                                         | We are inhibited when                                          | What inhibits                                                                                                                                                                                                                       | We are supported when                        | Opportunity to improve                                                              |
|-----------------------------------------------------------|----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|-------------------------------------------------------------------------------------|
| <b>Attention</b><br>How and where we focus our attention. | Attention is physiologically drawn, overwhelmed or fragmented. | <ul style="list-style-type: none"><li>• Constant context switching</li><li>• Many undifferentiated choices</li><li>• Fearful information</li><li>• No stopping cues (e.g. infinite scroll)</li><li>• Unnecessary movement</li></ul> | Enabled to bring more focus and mindfulness. |  |

Identify Opportunities to improve

# Humane Design Guide

<http://humanetech.com>

After analysis step, develop plan of action:

1. In what ways does your product/feature currently engage Human Sensitivities?
2. How might your product/feature support or elevate human sensitivities?
3. Action Statement

# GenderMag

<https://gendermag.org>

## Abby Jones<sup>1</sup>



### You can edit anything in blue print

- 28 years old
- Employed as an Accountant
- Lives in Cardiff, Wales

*Abby has always liked music. When she is on her way to work in the morning, she listens to music that spans a wide variety of styles. But when she arrives at work, she turns it off, and begins her day by scanning all her emails first to get an overall picture before answering any of them. (This extra pass takes time but seems worth it.) Some nights she exercises or stretches, and sometimes she likes to play computer puzzle games like Sudoku*

### Background and skills

Abby works as an accountant. She is comfortable with the technologies she uses regularly, but she just moved to this employer 1 week ago, and **their software systems are new to her**.

Abby says she's a "numbers person," but she has never taken any computer programming or IT systems classes. She **likes Math** and knows how to think with numbers. She writes and edits spreadsheet formulas in her work.

In her free time, she also **enjoys working with numbers and logic**. She especially likes working out puzzles and puzzle games, either on paper or on the computer

### Motivations and Attitudes

- **Motivations:** Abby uses technologies **to accomplish her tasks**. She learns new technologies if and when she needs to, but prefers to use methods she is **already familiar and comfortable with, to keep her focus** on the tasks she cares about.
- **Computer Self-Efficacy:** Abby has **low confidence about doing unfamiliar computing tasks**. If problems arise with her technology, she often **blames herself for these problems**. This affects whether and how she will persevere with a task if technology problems have arisen.
- **Attitude toward Risk:** Abby's life is a little complicated and she **rarely has spare time**. So she is **risk averse about using unfamiliar technologies that might need her to spend extra time** on them, even if the new features might be relevant. She instead performs tasks using familiar features, because they're more predictable about what she will get from them and how much time they will take.

### How Abby Works with Information and Learns:

- **Information Processing Style:** Abby tends towards a **comprehensive information processing style** when she needs to more information. So, instead of acting upon the first option that seems promising, she **gathers information comprehensively to try to form a complete understanding of the problem before trying to solve it**. Thus, her style is "burst-y"; first she reads a lot, then she acts on it in a batch of activity.
- **Learning: by Process vs. by Tinkering:** When learning new technology, Abby leans toward **process-oriented learning**, e.g., tutorials, step-by-step processes, wizards, online how-to videos, etc. She **doesn't particularly like learning by tinkering with software** (i.e., just trying out new features or commands to see what they do), but when she does tinker, it has positive effects on her understanding of the software.

<sup>1</sup>Abby represents users with motivations/attitudes and information/learning styles similar to hers. For data on females and males similar to and different from Abby, see <http://eusesconsortium.org/gender/gender.php>

# GenderMag

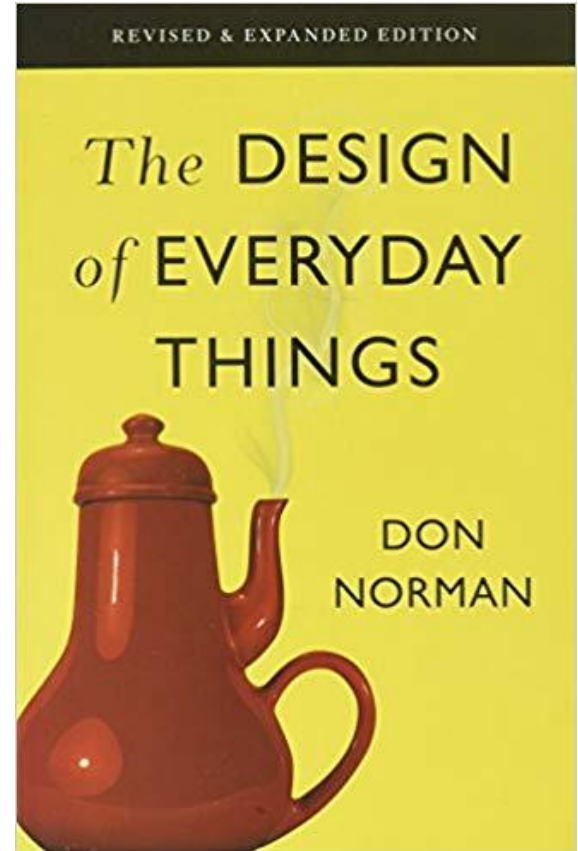
<https://gendermag.org>

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>1. Pick a persona. eg: Abby </li> <li>2. Pick a use case/scenario in your tool, eg:               <ul style="list-style-type: none"> <li>– in Book Store Navigator app...</li> <li>– “Find science fiction books”</li> </ul> </li> </ul>                                                                                                              | <ul style="list-style-type: none"> <li>3a-b. Pick a Subgoal for that scenario. eg: </li> </ul> <p>Subgoal #1: “See bookstore map”.</p> <p>Q: Will <b>Abby</b> have formed this sub-goal...?</p> <ul style="list-style-type: none"> <li>• Yes/no/<i>maybe</i>. Why? Consider <i>Abby's Motivations...</i></li> </ul>  |
| <ul style="list-style-type: none"> <li>3c-d. Pick an Action for that subgoal. </li> </ul> <p>Action #1: “Tap ‘Browse Off’”:</p> <p>– Q1. Will <b>Abby</b> <u>know what to do</u>?</p> <ul style="list-style-type: none"> <li>• Yes/no/<i>maybe</i>. Why? Consider Abby's, ... <i>Tinkering</i></li> </ul>  <p>→ First answer Q1.<br/>After answering it, <u>then</u> perform the action.</p> | <p>– 3e. Q2. If she performs the action, producing </p>  <p>will <b>Abby</b> <u>see progress</u> toward the subgoal?</p> <ul style="list-style-type: none"> <li>• Yes/no/<i>maybe</i>. Why? Consider Abby's <i>Self-Efficacy</i> &amp; ...</li> </ul>                                                                |

# User Centered Design

User-centered design tries to optimize the product around how **users can, want, or need to use the product**, rather than forcing the users to change their behavior to **accommodate the product**.

-Wikipedia

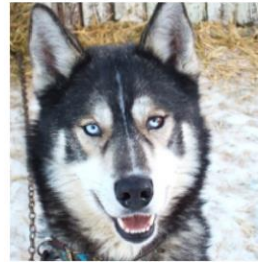


# Agile

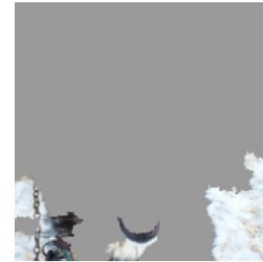


**2. Does my software amplify positive or negative behavior for users and society at large?**

# Dog vs Wolf



(a) Husky classified as wolf



(b) Explanation

**Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.**

|                             | Before       | After        |
|-----------------------------|--------------|--------------|
| Trusted the bad model       | 10 out of 27 | 3 out of 27  |
| Snow as a potential feature | 12 out of 27 | 25 out of 27 |

# Local Interpretable Model-Agnostic Explanations (LIME)

<https://github.com/marcotcr/lime>

Prediction probabilities



atheism

christian

Posting  
0.15  
Host  
0.14  
NNTP  
0.11  
edu  
0.04  
have  
0.01  
There  
0.01

## Text with highlighted words

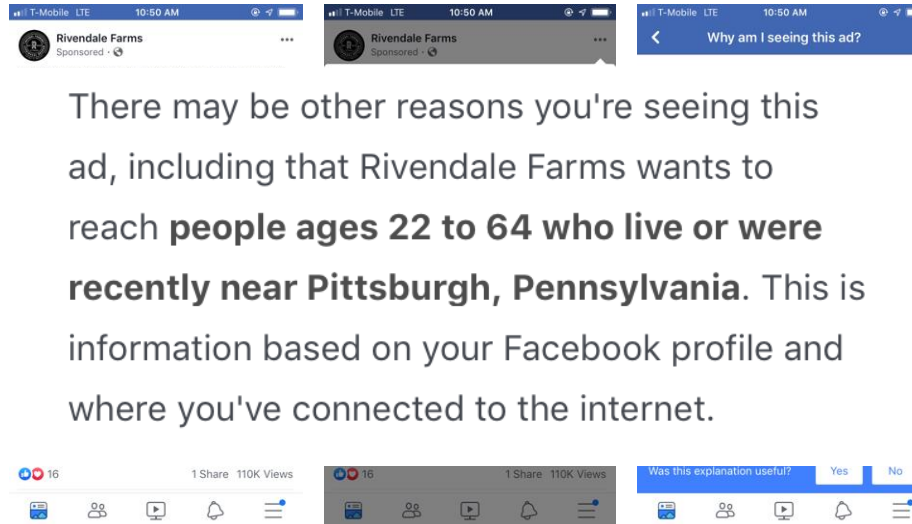
From: johnchad@triton.unm.edu (jchadwic)  
Subject: Another request for Darwin Fish  
Organization: University of New Mexico, Albuquerque  
Lines: 11  
NNTP-Posting-Host: triton.unm.edu

Hello Gang,

There have been some notes recently asking where to obtain the DARWIN fish.

This is the same question I have and I have not seen an answer on the net. If anyone has a contact please post on the net or email me.

# Explain “why” to customers





@dovneon

## What Instagram removing likes may mean for influencers and our self-esteem

SCIENCE & TECH - FEATURE

The decision could have a positive impact on the way people use the platform, but harm those trying to use it professionally

# Anil Dash on how to prevent abuse

[http://anildash.com/2011/07/20/if\\_your\\_websites\\_full\\_of\\_assholes\\_its\\_your\\_fault-2/](http://anildash.com/2011/07/20/if_your_websites_full_of_assholes_its_your_fault-2/)

You should have real humans dedicated to monitoring and responding to your community.

You should have community policies about what is and isn't acceptable behavior.

Your site should have accountable identities.

You should have the technology to easily identify and stop bad behaviors.

You should make a budget that supports having a good community, or you should find another line of work.



tests **passing** codecov **97%** pypi **v0.2.2** conda-forge **v0.2.2**

[Read more about deon on the project homepage](#)

---

## An ethics checklist for data scientists

---

`deon` is a command line tool that allows you to easily add an ethics checklist to your data science projects. We support creating a new, standalone checklist file or appending a checklist to an existing analysis in [many common formats](#).

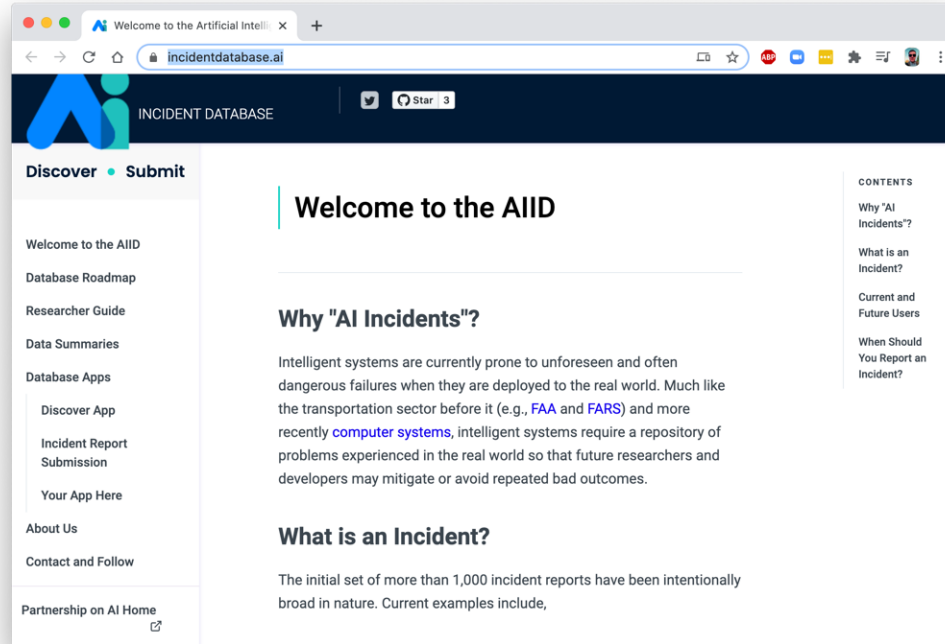
---

δέον • (déon) [n.] (*Ancient Greek*) [wiktionary](#)

Duty; that which is binding, needful, right, proper.

---

# AI Incident Database



**3. Will my software's  
quality impact the  
humanity of others?**

# Quality has long been considered

## Quality attributes [\[ edit \]](#)

Notable quality attributes include:

- [accessibility](#)
- [accountability](#)
- [accuracy](#)
- [adaptability](#)
- [administrability](#)
- [affordability](#)
- [agility](#) [Toll] (see Common Subsets below)
- [auditability](#)
- [autonomy](#) [Erl]
- [availability](#)
- [compatibility](#)
- [composability](#) [Erl]
- [configurability](#)
- [correctness](#)
- [credibility](#)
- [customizability](#)
- [debugability](#)
- [degradability](#)
- [determinability](#)
- [demonstrability](#)
- [dependability](#)
- [deployability](#)
- [discoverability](#) [Erl]
- [distributability](#)
- [durability](#)
- [effectiveness](#)
- [efficiency](#)
- [evolvability](#)
- [extensibility](#)
- [failure transparency](#)
- [fault-tolerance](#)
- [fidelity](#)
- [flexibility](#)
- [inspectability](#)
- [installability](#)
- [integrity](#)
- [interchangeability](#)
- [interoperability](#) [Erl]
- [learnability](#)
- [localizability](#)
- [maintainability](#)
- [manageability](#)
- [mobility](#)
- [modifiability](#)
- [modularity](#)
- [observability](#)
- [operability](#)
- [orthogonality](#)
- [portability](#)
- [precision](#)
- [predictability](#)
- [process capabilities](#)
- [producibility](#)
- [provability](#)
- [recoverability](#)
- [relevance](#)
- [reliability](#)
- [repeatability](#)
- [reproducibility](#)
- [resilience](#)
- [responsiveness](#)
- [reusability](#) [Erl]
- [robustness](#)
- [safety](#)
- [scalability](#)
- [seamlessness](#)
- [self-sustainability](#)
- [serviceability](#) (a.k.a. supportability)
- [securability](#)
- [simplicity](#)
- [stability](#)
- [standards compliance](#)
- [survivability](#)
- [sustainability](#)
- [tailorability](#)
- [testability](#)
- [timeliness](#)
- [traceability](#)
- [transparency](#)
- [ubiquity](#)
- [understandability](#)
- [upgradability](#)
- [vulnerability](#)
- [usability](#)

# Engineering ethics.

Ethics applies and is formalized in many professional fields: medical, legal, business, and engineering.

The first codes of engineering ethics were formally adopted by American engineering societies in 1912-1914. In 1946 the National Society of Professional Engineers (NSPE) adopted their first formal Canons of Ethics.

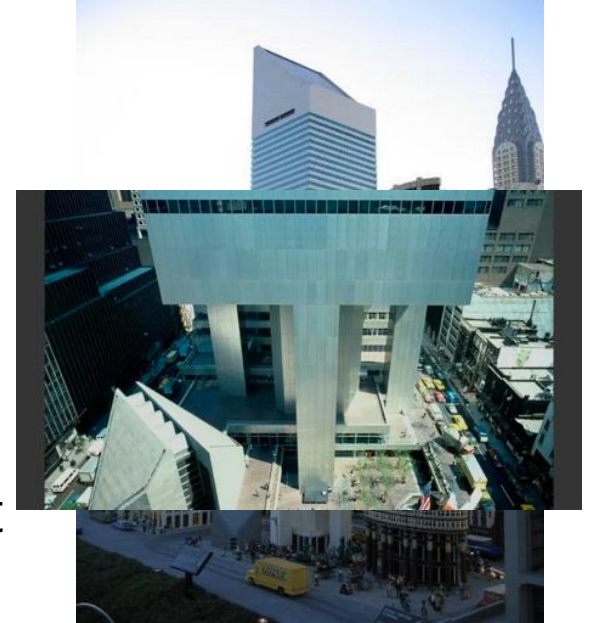
# “hold paramount safety, health and welfare of the public”

Citigroup Center, Designed by Structural engineer William LeMessurier

Followed calculations required by building codes

Civil Engineering student Diane Hartley realized there was a problem

Tests showed that winds needed to bring it down would happen every 55 years



# Professional Ethics

Professional ethics encompass the personal, and corporate standards of behavior expected by professionals.

First three “professions”

- Divinity,
- Law
- Medicine



# Medicine - Intrinsic

## Hippocratic Oath

~450BC

“Do no Harm”



# Law -Extrinsic

Bar regulates behavior

Oath to follow rules

Malpractice



# Legal Malpractice

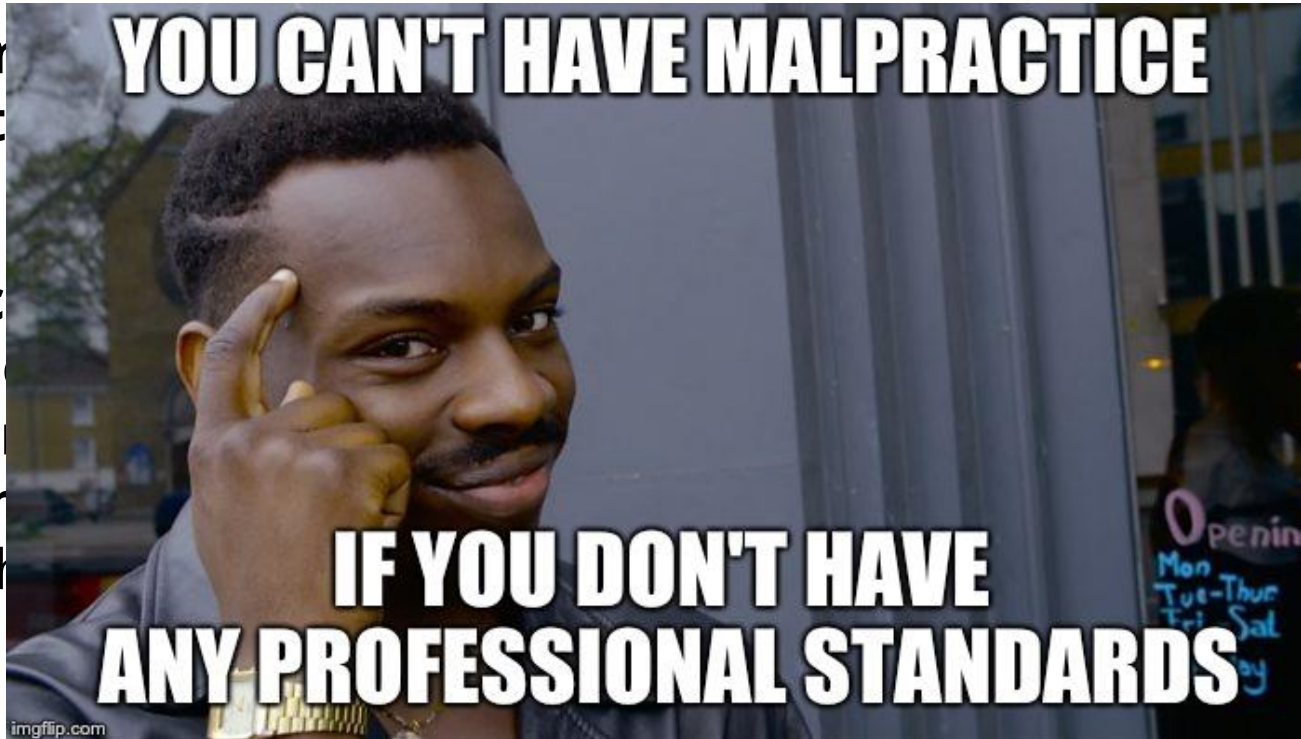
Not every mistake is legal malpractice. For malpractice to exist:

Attorney must handle a case inappropriately  
due to negligence or with intent to harm  
And cause damages to a client

# Malpractice vs. Negligence

**Negligence**  
prudent

**Malpractice**  
negligence  
lawyer or  
**set by the**  
causing harm



ply

ssional  
doctor,  
**standards**  
ntly

# DISCUSSION: What should we do going forward?

# Professional Engineers

What {is / could be} the role of **professional engineers** in software?



By ----PCStuff 03:47, 31 July 2006 (UTC), CC BY-SA 2.5,  
<https://commons.wikimedia.org/w/index.php?curid=10340855>

# Professional Engineers

What {is / could be} the role of **professional engineers** in software?



Maybe coming to CMU soon?



By ----PCStuff 03:47, 31 July 2006 (UTC), CC BY-SA 2.5,  
<https://commons.wikimedia.org/w/index.php?curid=10340855>

# Will software quality impact human flourishing?

Most traditional emphasis of “engineering ethics”

What can we learn from other professions?

Should software have “Professional Engineers”?

How do we define “safety critical systems”?

How much testing is enough? How can we convince others to do that much testing?

These questions are the **start** of the **conversation**, but as technology evolves, we must be **vigilant** to ensure we are promoting human flourishing

Interested in more?

17-200 Ethics and  
Policy in computing

# Three questions to promote human flourishing

1. Does my software respect the **humanity** of the **users**?
2. Does my software **amplify positive** behavior, or **negative** behavior for users and society at large?
3. Will my software's **quality** impact the **humanity** of others?