

# Ethics

Michael Hilton



Software and Societal  
Systems Department

# Administrivia

**Due Today: Project 4 checkpoint**

**Final Project 4 – Tuesday April 11**

**No class Thursday April 13 – Carnival**

**Project 5 is live: <https://cmu-313.github.io/projects/P5/>**

**Let me know if you are concerned about finding an internship, and would be interested in a class over the summer.**

# Learning goals

- Awareness of ethical issues in software engineering
- Reflection on decision making
- Questions to ask when evaluating the ethics of software
- Starting points to dig deeper

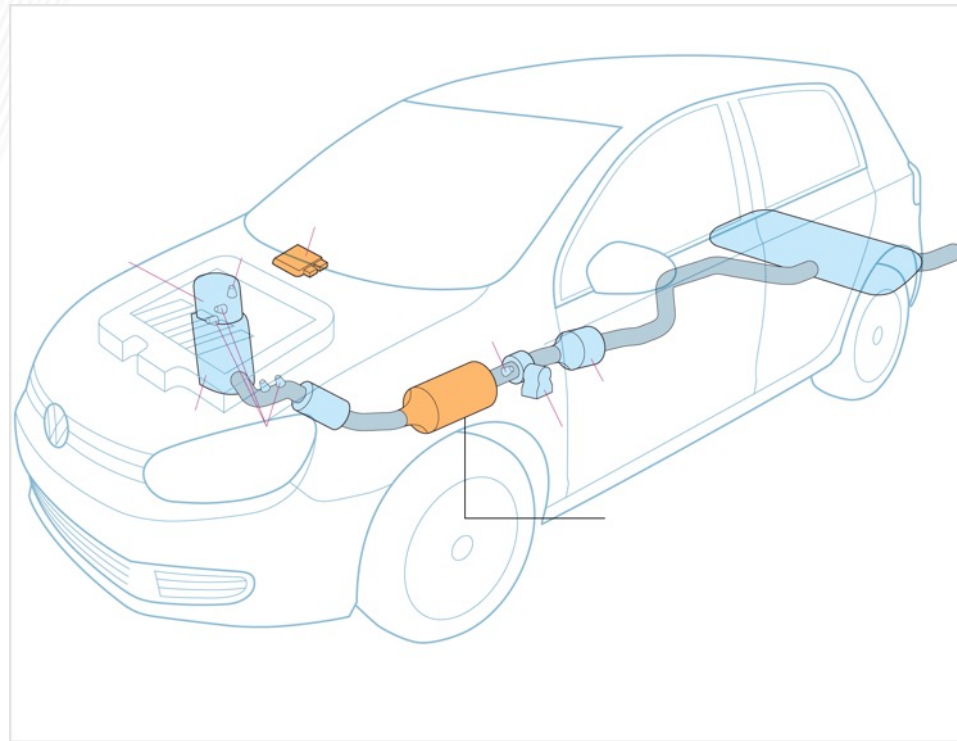


# Ethics



# 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](#) [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](#) [t](#) [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](#) [Twitter](#) [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



Currently, the AI portrait generator has been trained mostly on portraits of people of European ethnicity. We're planning to expand our dataset and fix this in the future. At the time of conceptualizing this AI, authors were not certain it would turn out to work at all. This is close to state of the art in AI at the moment.

Sorry for the bias in the meanwhile. Have fun!

324 Retweets 65 Quote Tweets 1,243 Likes



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

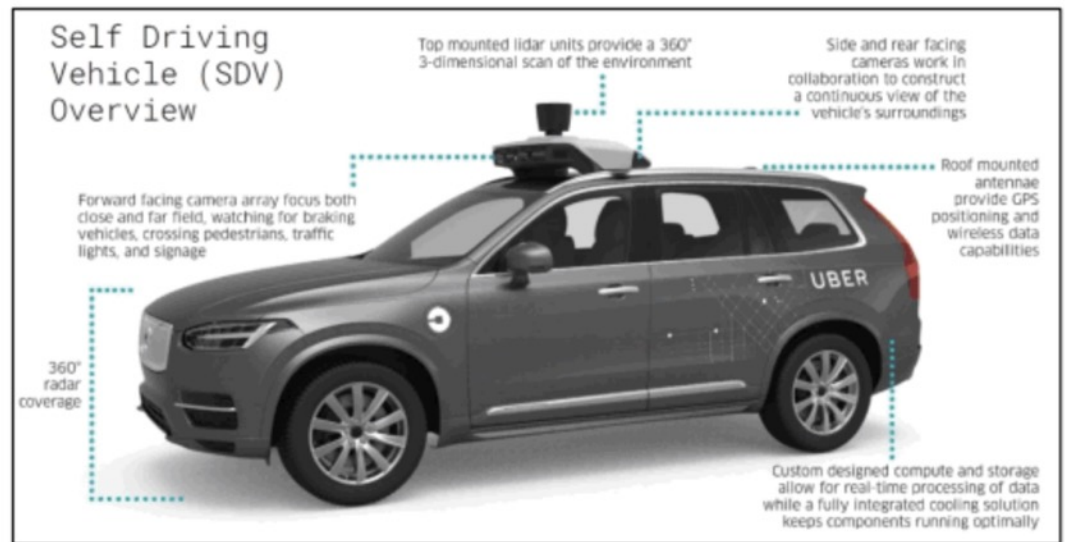
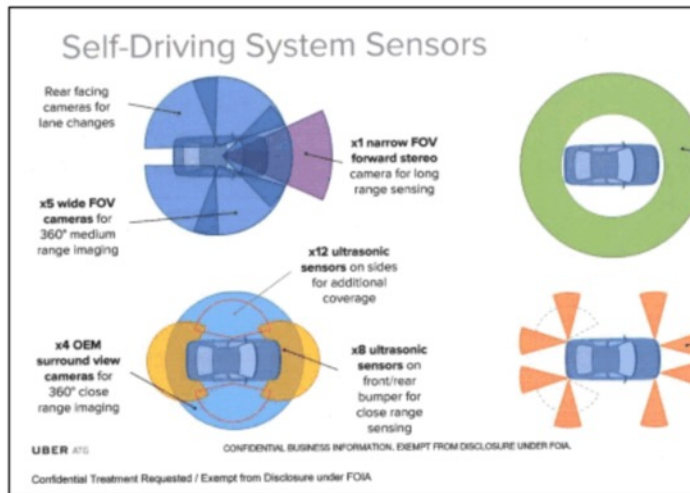
The system had several serious software flaws, the NTSB said.



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

25  
Comments

1131  
Shares



# xing.com search for “Brand Strategist”

Search query	Work experience	Education experience	Profile views	Candidate	Xing ranking
Brand Strategist	146	57	12992	male	1
Brand Strategist	327	0	4715	female	2
Brand Strategist	502	74	6978	male	3
Brand Strategist	444	56	1504	female	4
Brand Strategist	139	25	63	male	5
Brand Strategist	110	65	3479	female	6
Brand Strategist	12	73	846	male	7
Brand Strategist	99	41	3019	male	8
Brand Strategist	42	51	1359	female	9
Brand Strategist	220	102	17186	female	10

Lahoti, Preethi, Krishna P. Gummadi, and Gerhard Weikum. “iFair: Learning Individually Fair Data Representations for Algorithmic Decision Making.” 2019 IEEE 35th International Conference on Data Engineering (ICDE) (2019)

# Twitter cropping photos

 **Tony "Abolish (Pol)ICE" Arcieri**   
@bascule

Trying a horrible experiment...

Which will the Twitter algorithm pick: Mitch McConnell or Barack Obama?



6:05 PM · Sep 19, 2020 · Twitter Web App

**64.7K** Retweets   **16.3K** Quote Tweets   **198.6K** Likes





# Twitter cropping photos

**Tony "Abolish (Pol)ICE" Arcieri** @bascule

Trying a horrible experiment...

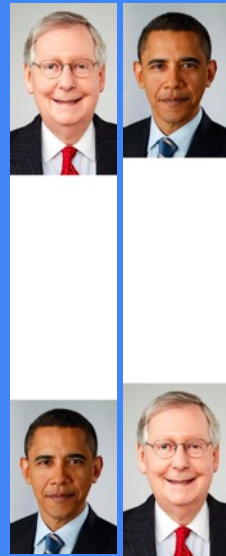
Which will the Twitter algorithm pick: Mitch McConnell or Barack Obama?



6:05 PM · Sep 19, 2020 · Twitter Web App

**64.7K** Retweets **16.3K** Quote Tweets **198.6K** Likes

Reply icons: Comment, Retweet, Like, Share



**TheArtGun COMMS OPEN** @TheArtGun

Replying to @bascule

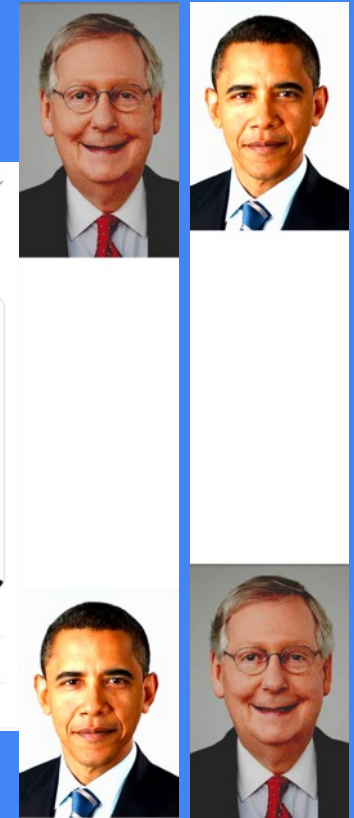
What if we adjust the contrast



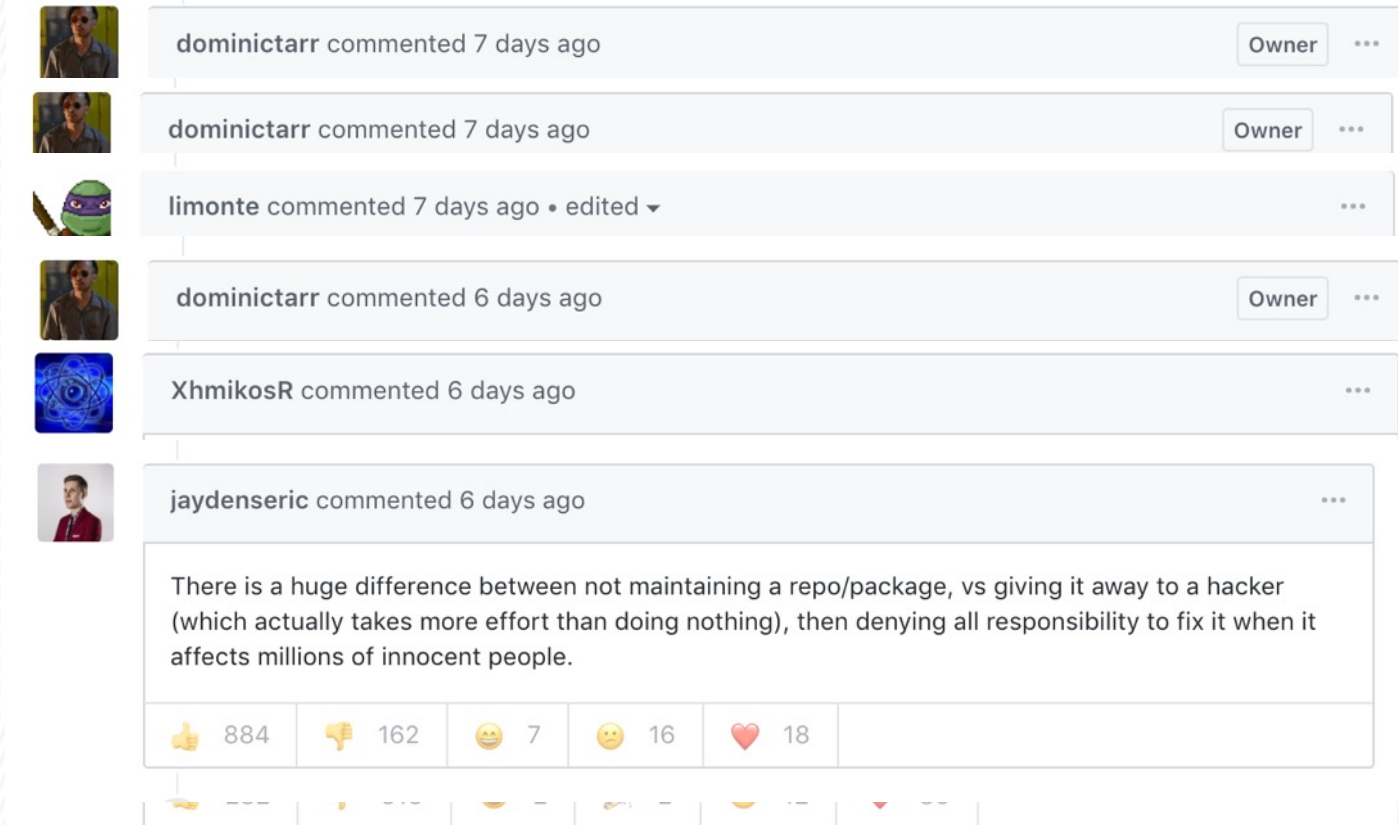
10:36 PM · Sep 19, 2020 · Twitter Web App

**35** Retweets **5** Quote Tweets **102** Likes

Reply icons: Comment, Retweet, Like, Share



# Open Source Maintainers



The screenshot shows a series of GitHub comments. The first comment is from dominictarr, posted 7 days ago, with an 'Owner' label and a three-dot menu. The second comment is also from dominictarr, posted 7 days ago, with an 'Owner' label and a three-dot menu. The third comment is from limonte, posted 7 days ago, with an 'edited' dropdown and a three-dot menu. The fourth comment is from dominictarr, posted 6 days ago, with an 'Owner' label and a three-dot menu. The fifth comment is from XhmikosR, posted 6 days ago, with a three-dot menu. The sixth comment is from jaydenseric, posted 6 days ago, with a three-dot menu. Below the sixth comment is a text block: 'There is a huge difference between not maintaining a repo/package, vs giving it away to a hacker (which actually takes more effort than doing nothing), then denying all responsibility to fix it when it affects millions of innocent people.' Below the text block are reaction counts: 884 thumbs up, 162 thumbs down, 7 neutral faces, 16 sad faces, and 18 hearts.

dominictarr commented 7 days ago Owner ...

dominictarr commented 7 days ago Owner ...

limonte commented 7 days ago • edited ▾ ...

dominictarr commented 6 days ago Owner ...

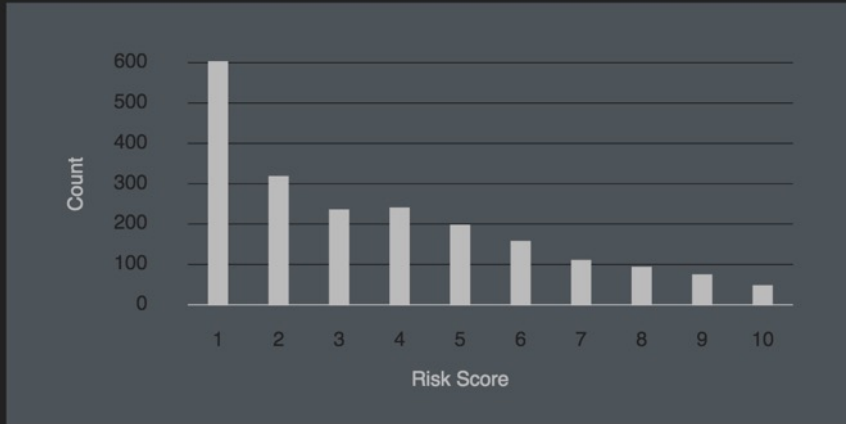
XhmikosR commented 6 days ago ...

jaydenseric commented 6 days ago ...

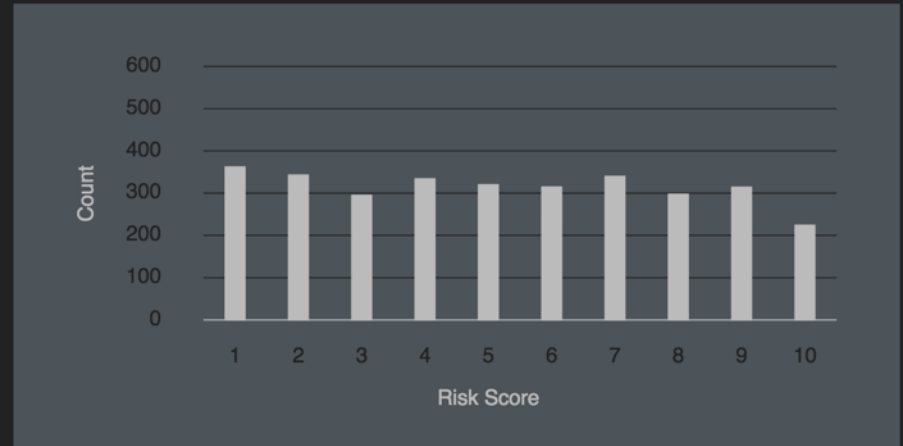
There is a huge difference between not maintaining a repo/package, vs giving it away to a hacker (which actually takes more effort than doing nothing), then denying all responsibility to fix it when it affects millions of innocent people.

👍 884 👎 162 😐 7 😞 16 ❤️ 18

### White Defendants' Risk Scores



### Black 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.)

## 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

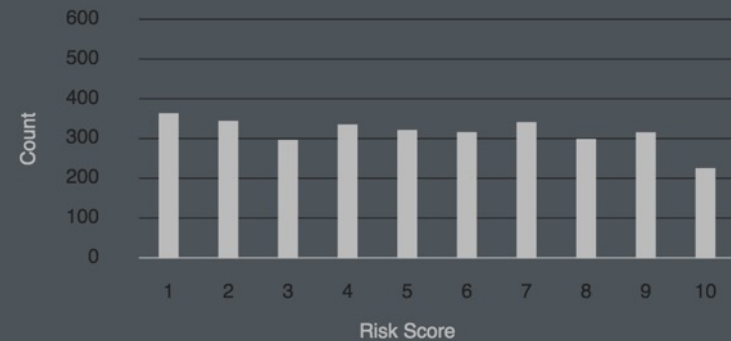
Access to health care

Receiving parole

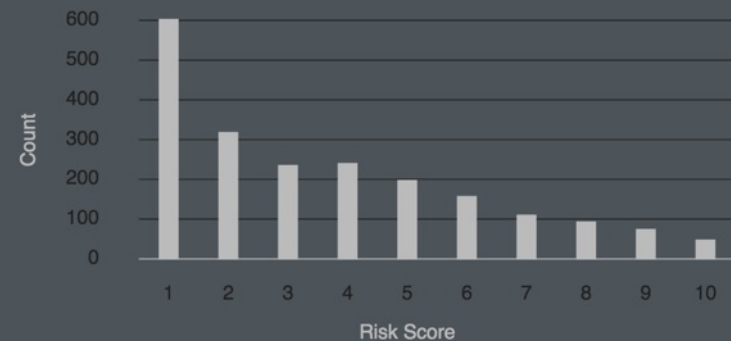
Possibility of Bail

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

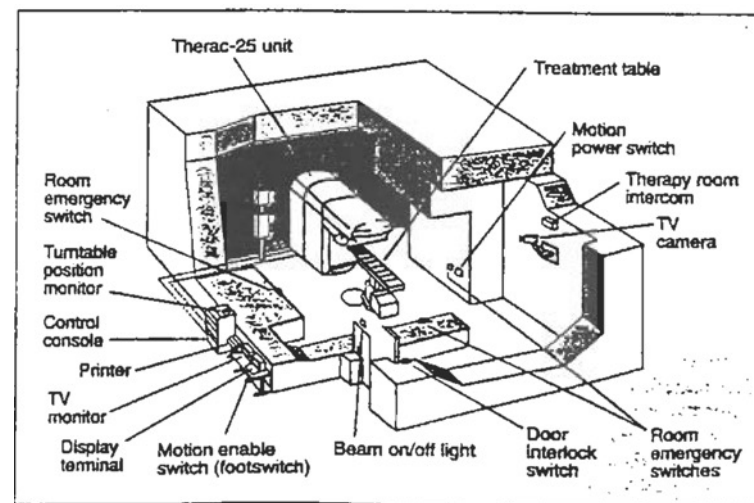
Traced to:

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
TREATMENT MODE: FIX
BEAM TYPE: E ENERGY (KeV): 10

UNIT RATE/MINUTE ACTUAL PRESCRIBED
MONITOR UNITS 200.000000 200.000000
TIME (MIN) 0.270000 0.270000

GANTRY 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
WEDGE NUMBER 1.000000 1.000000 VERIFIED
ACCESSORY NUMBER 0.000000 0.000000 VERIFIED

DATE: 2012-04-16 SYSTEM: BEAM READY OP. MODE: TREAT AUTO
TIME: 11:49:59 TREAT: TREAT PAUSE 3-DAY 173777
OPR ID: 033-1f3sp REASON: OPERATOR COMMAND: █
```

Make a contribution

Subscribe Find

BUINESS DAY

US edition

News

4,331 views | Oct 17, 2018, 06:13pm

n

The

DEV

GUES

Eng  
Opin

# We Need To Work Harder To Make Software Engineering More Ethical

it ethics

☰

Software body

OLGA V. MACK



Jessica Baron Contributor  
Consumer Tech

I write about the ethics of science and technology.

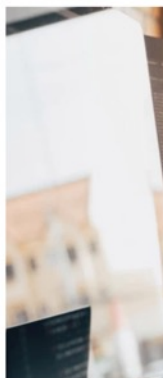
Ho

FEJ

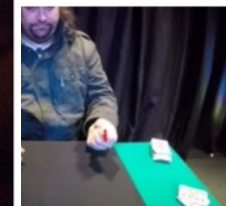
S

W

et



T READ



to fool AI with magic

patch the software, but you can't patch a person if you, you know, damage someone's reputation." Sam Hodgson for The New York Times



# 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> 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.	 High Low
<b>Attention</b> 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> 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> 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> 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> 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> 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.	

# 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>

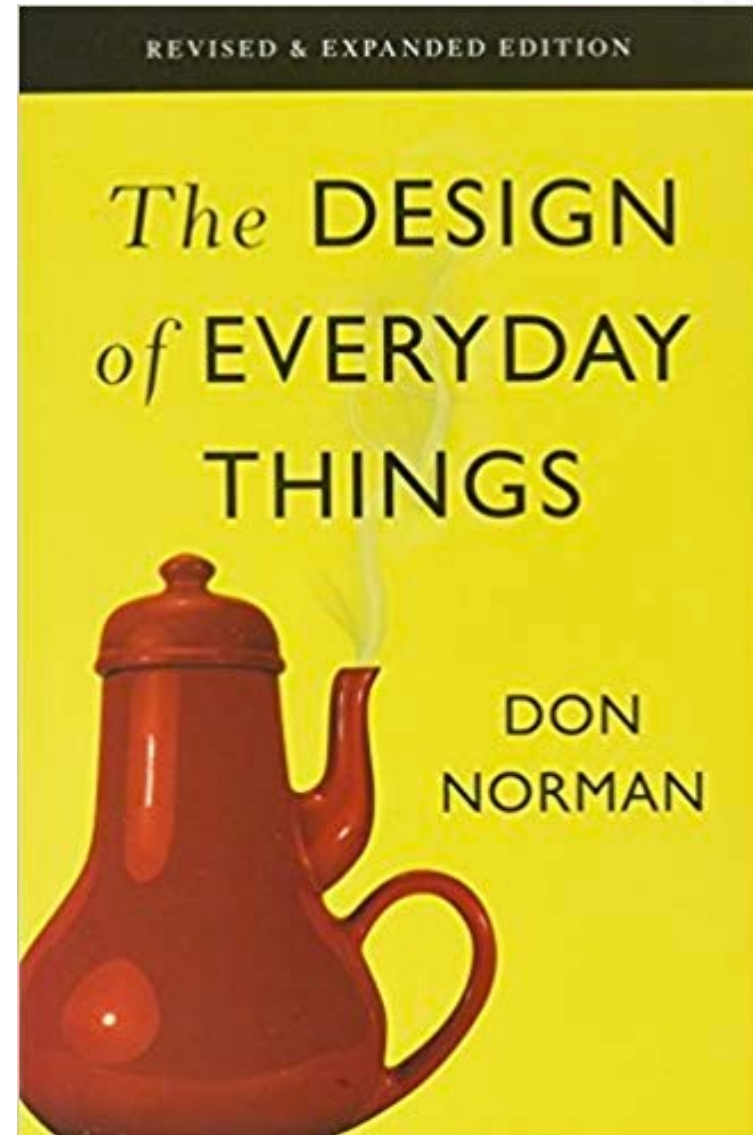
<ul style="list-style-type: none"> <li>• 1. Pick a persona. eg: Abby </li> <li>• 2. Pick a use case/scenario in your tool, eg:             <ul style="list-style-type: none"> <li>– in Book Store Navigator app...</li> <li>– “Find science fiction books”</li> </ul> </li> </ul> 	<ul style="list-style-type: 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> <ul style="list-style-type: none"> <li>– Q1. Will <b>Abby</b> know what to do?             <ul style="list-style-type: none"> <li>• Yes/no/<i>maybe</i>. Why? Consider <i>Abby's, ... Tinkering</i></li> </ul> </li> </ul>  <p>→ <b>First answer Q1.</b>  <b>After answering it, then perform the action.</b></p>	<ul style="list-style-type: none"> <li>– 3e. Q2. If she performs the action, producing </li> </ul>  <p>will <b>Abby</b> see progress toward the subgoal?</p> <ul style="list-style-type: none"> <li>• Yes/no/<i>maybe</i>. Why? Consider <i>Abby's Self-Efficacy &amp; ...</i></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

User C

Agile c



**Manifesto for Agile Software Development**

We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

Individuals and interactions over processes and tools

**Customer collaboration over contract negotiation**

Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

Kent Beck	James Grenning	Robert C. Martin
Mike Beedle	Jim Highsmith	Steve Mellor
Arie van Bennekum	Andrew Hunt	Ken Schwaber
Alistair Cockburn	Ron Jeffries	Jeff Sutherland
Ward Cunningham	Jon Kern	Dave Thomas
Martin Fowler	Brian Marick	

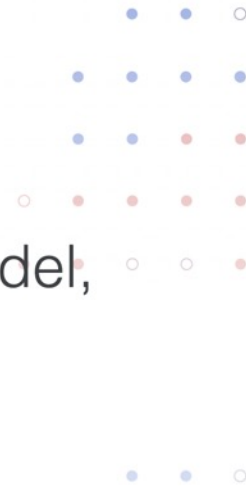
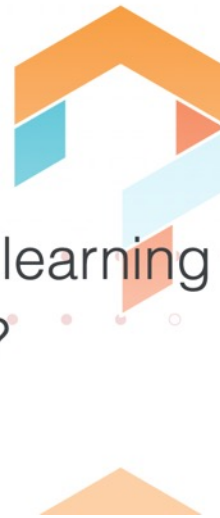
**2. Does my software amplify positive or negative behavior for users and society at large?**

# What if...

<https://pair-code.github.io/what-if-tool/>

## What If...

you could inspect a machine learning model,  
with minimal coding required?





# What if...

<https://pair-code.github.io/what-if-tool/>

What-If Tool demo - binary classifier for predicting salary of over \$50k - UCI census income dataset

Partial dependence plots   Compute distance   Show nearest different classification: L1   L2 ⓘ

PERFORMANCE + FAIRNESS   **DATAPoint EDITOR**   FEATURES

Binning | X-Axis: age   Co... Binning | Y-Axis: marital-stat...   C... Color By: Inference

17-23   24-29   30-35   36-42   43-48

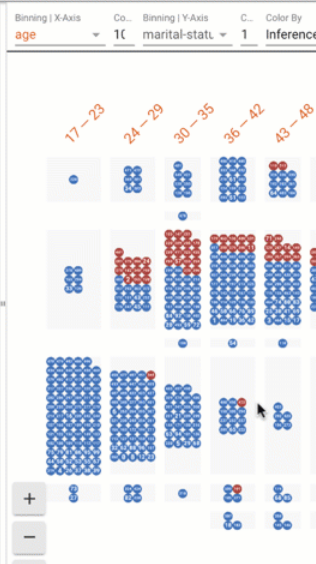
Select a datapoint to begin exploring features and values. →

Clicking on a datapoint in the visualization will load all the features and values associated with that example. Here are some of the things you can do:

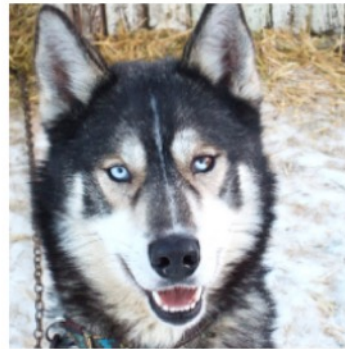
- **Edit features and values and rerun inference** to see how your model performs.
- **Compute Distance**: Select an example to be an anchor and create a new L1 or L2 distance feature for all loaded examples.
- **Closest Counterfactuals**: For classification models, find the closest example with a different classification using L1 or L2 distance.
- **Partial Dependence Plots**: For a selected example, explore plots for every feature that show the change in inference results across different valid values for that feature.

Use the Performance + Fairness tab to investigate model performance across your dataset.

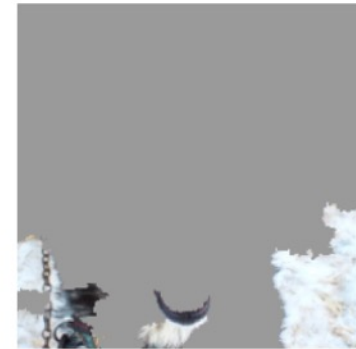
Use the Features tab to view statistics about your dataset.

The screenshot shows the 'What-If Tool' interface. At the top, there's an orange header with the title 'What-If Tool demo - binary classifier for predicting salary of over \$50k - UCI census income dataset'. Below the header, there are several tabs: 'PERFORMANCE + FAIRNESS', 'DATAPoint EDITOR' (which is currently selected), and 'FEATURES'. The 'DATAPoint EDITOR' tab contains a large area with instructions and a list of actions. To the right of the text is a grid of visualizations. The top row of the grid shows five columns representing age bins: '17-23', '24-29', '30-35', '36-42', and '43-48'. Each column contains a plot of data points (blue and red dots) representing different marital status categories. The plots show how the distribution of marital status changes across different age groups. There are also '+' and '-' buttons at the bottom left of the grid.

# 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)

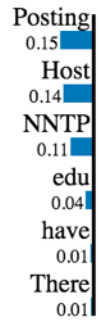
<https://github.com/marcotcr/lime>

Prediction probabilities



atheism

christian



## 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,

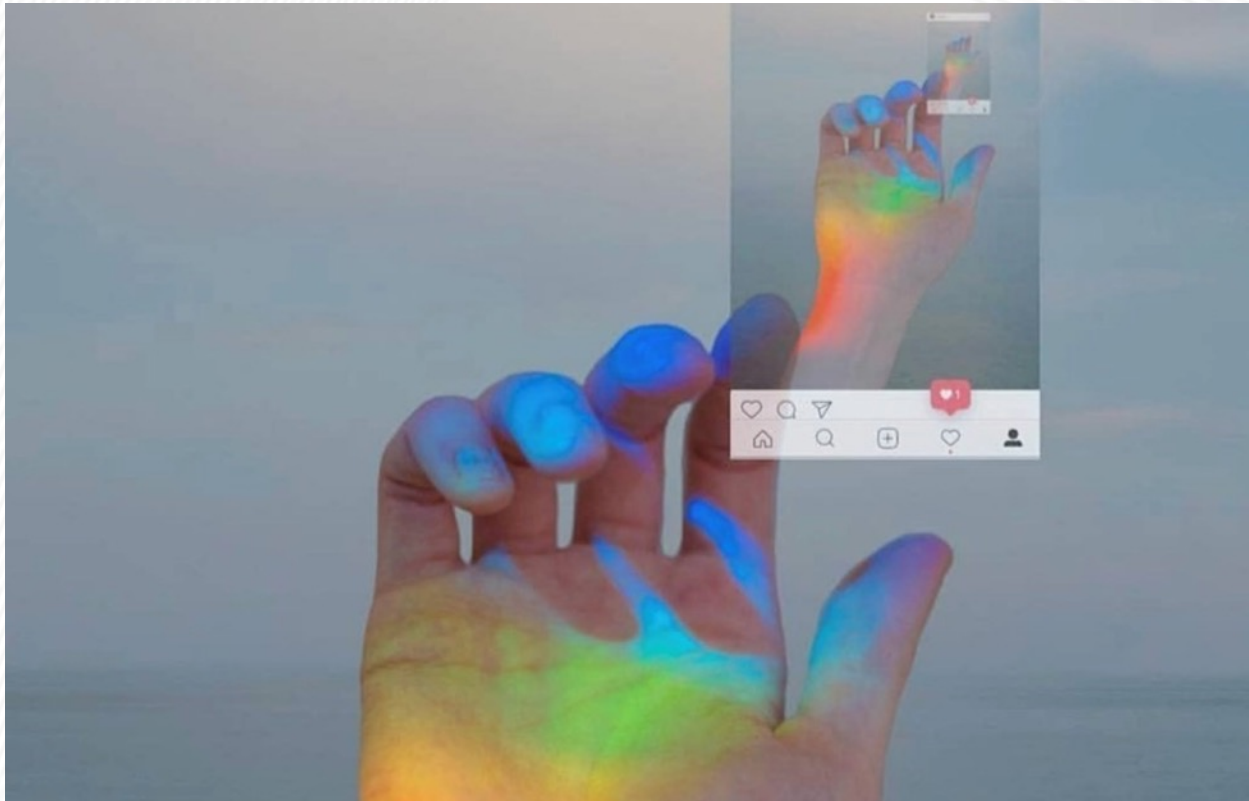
There have been some notes recently asking where to obtain the DARWIN fish.

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.

# Deon <https://github.com/drivendataorg/deon>



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

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

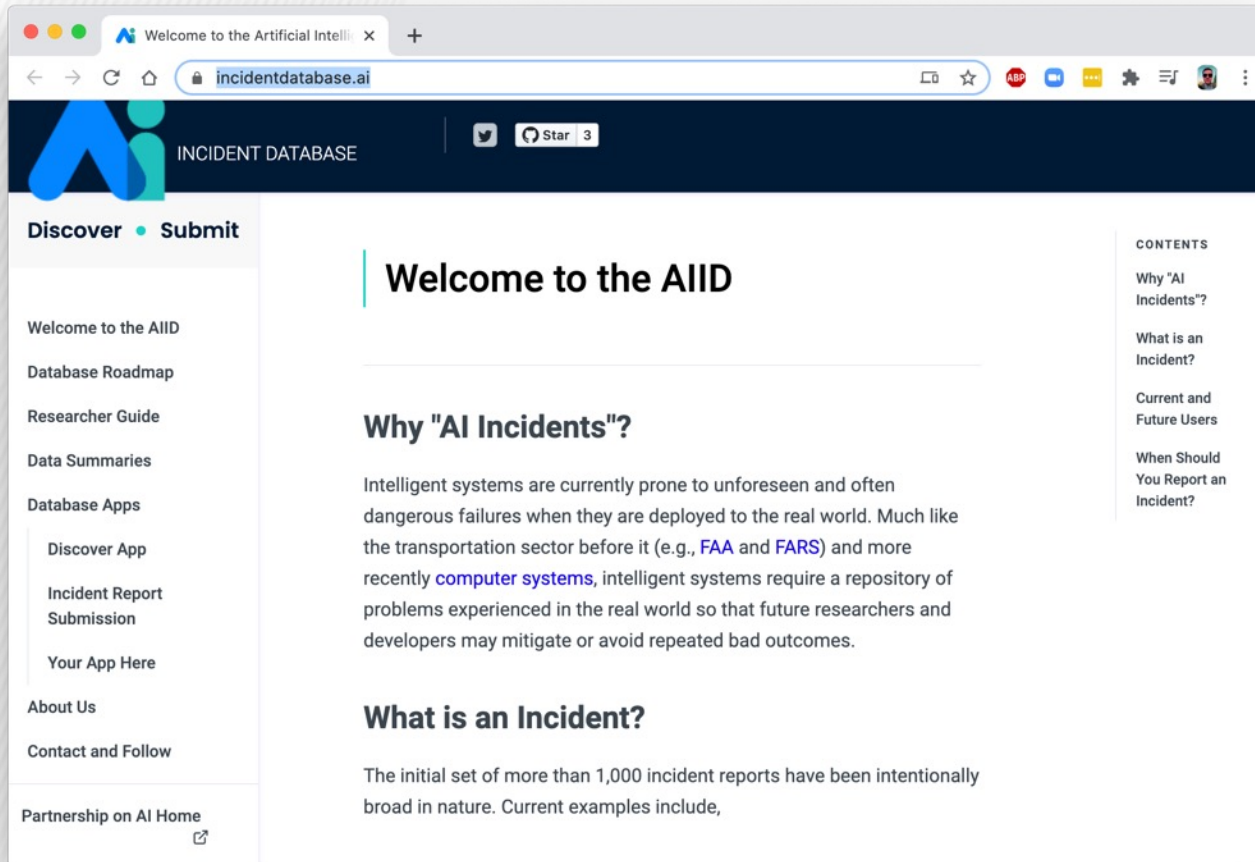
## An ethics checklist for data scientists

`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*) [wikitionary](#)

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

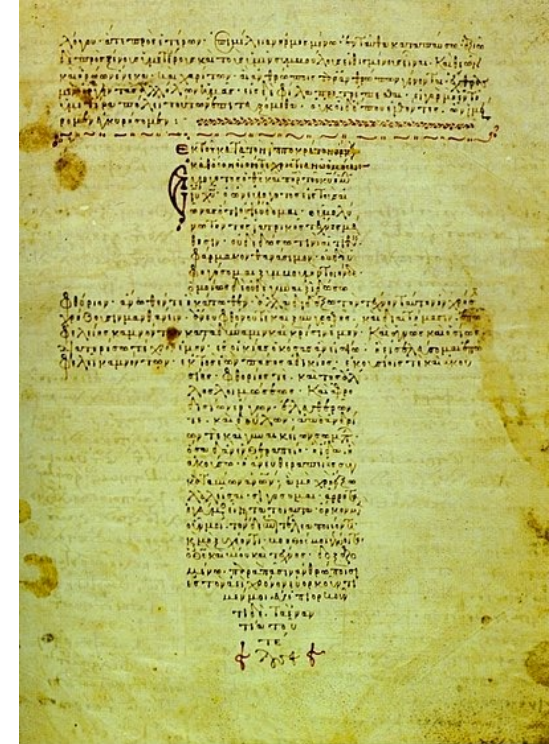


# Medicine - Intrinsic

# 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

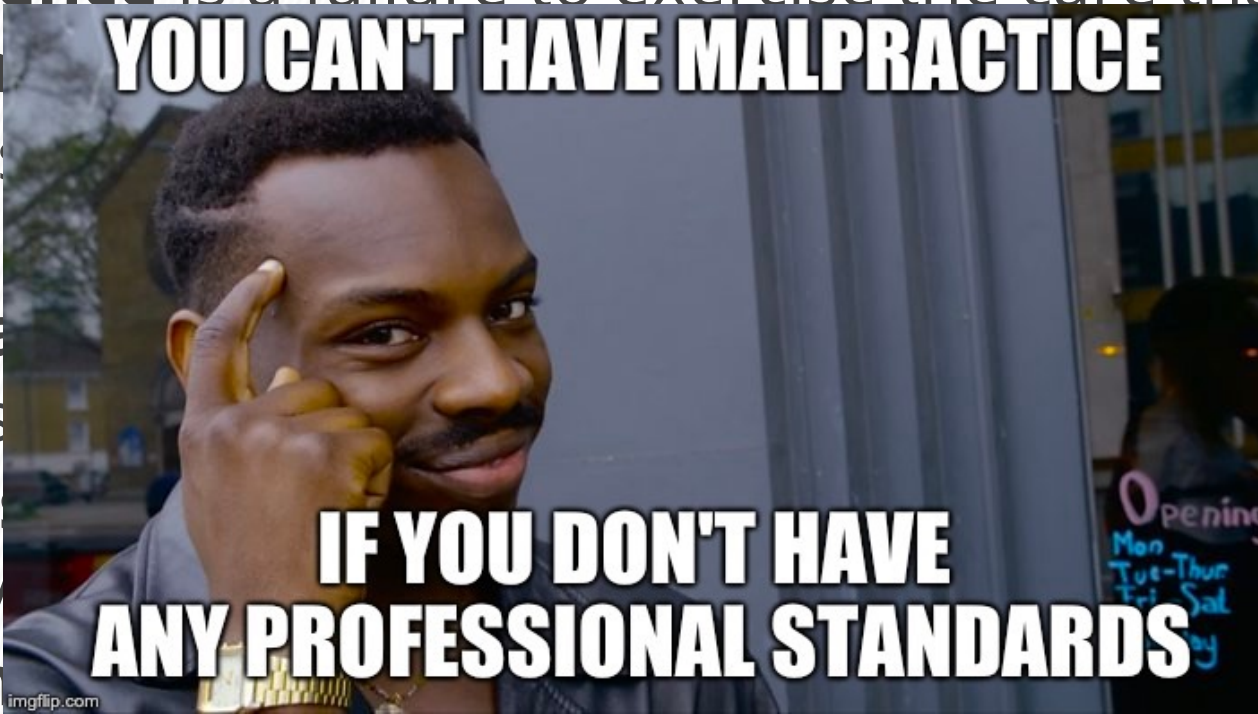


# Malpractice vs. Negligence

**Negligence** is a failure to exercise the care that a reasonable person would exercise in like circumstances.

**Malpractice** is a failure to exercise the care that a reasonable professional (called a "professional standard") fails to provide to a patient by the government or a professional.

causing harm to the plaintiff.



imgflip.com

**DISCUSSION: What should we do going forward?**

# Bioengineering Ethics:

- Respect for Autonomy
- Beneficence
- Nonmaleficence
- Justice

# Professional Engineers

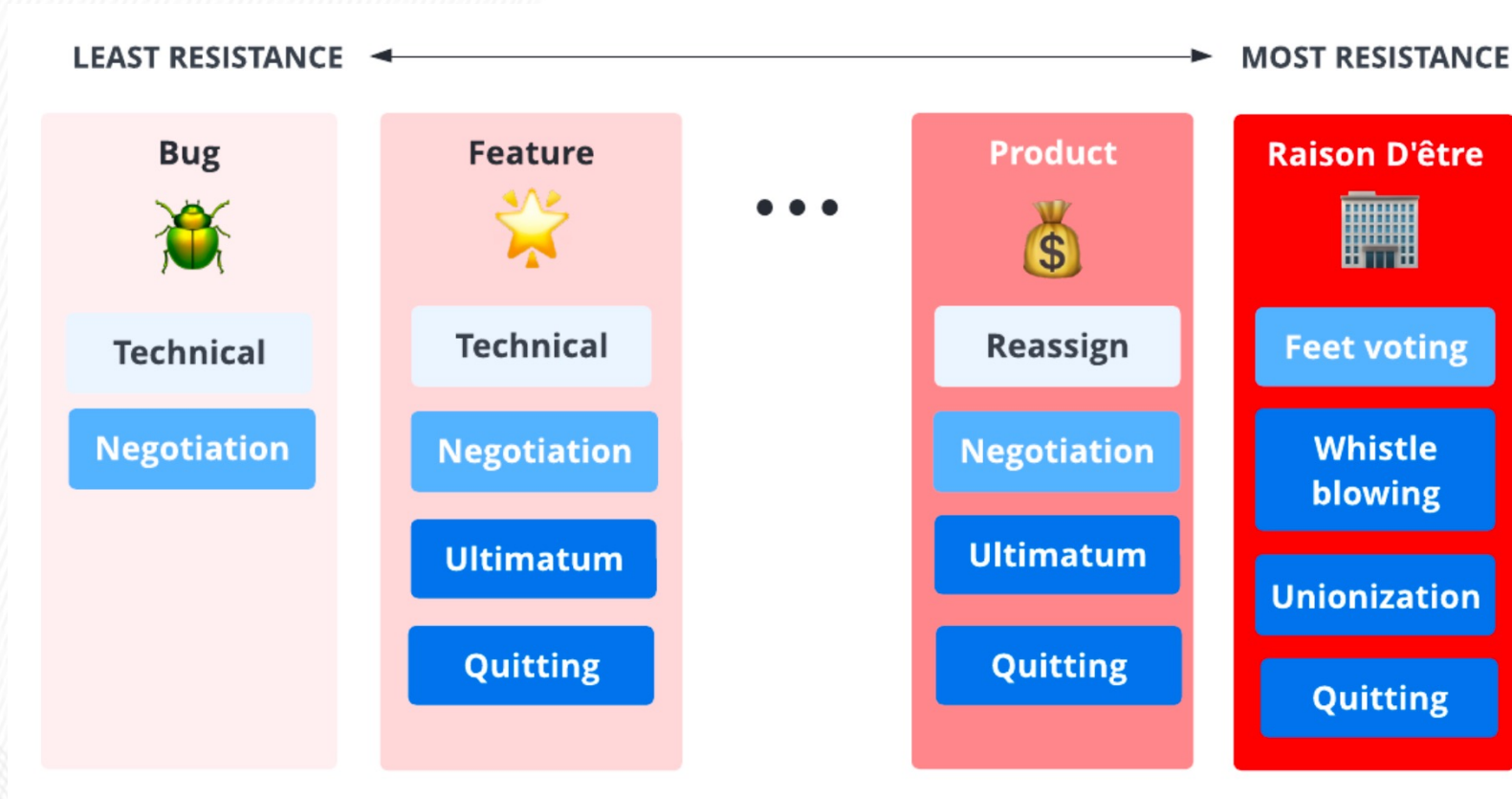
What {is / could be} the role of **professional engineers** in software?



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



# Different scope of concerns addressed differently



## 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

# 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?