

Ethics

Michael Coblenz

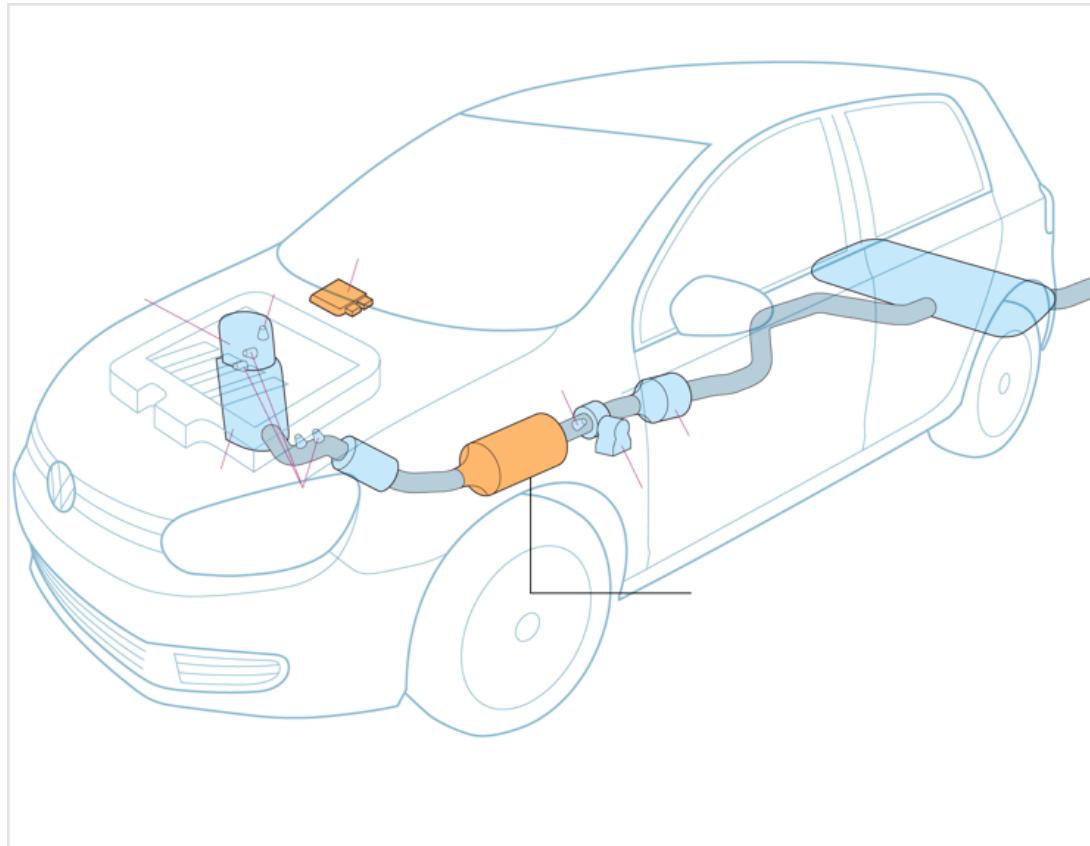
(slide credit: Michael Hilton at Carnegie Mellon)

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.

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>

Activity:
(Un)Ethical situations

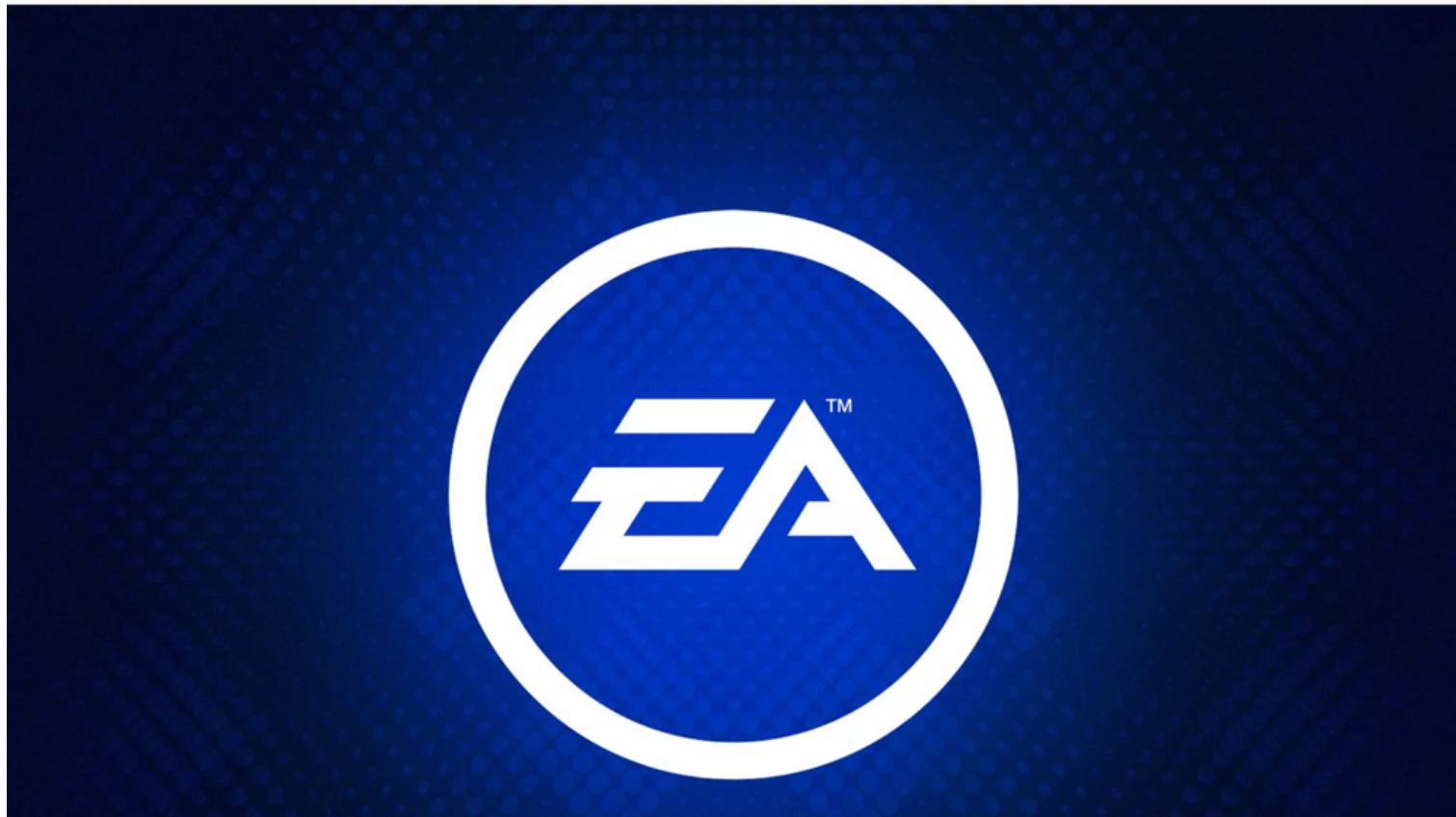
EA calls its loot boxes ‘surprise mechanics,’ says they’re used ethically

‘People like surprises,’ executive tells UK Parliament

By Ana Diaz | @AnaLikesPikachu | Jun 21, 2019, 9:10am EDT



 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

   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 | Nov 27, 2018, 6:10pm EST

f   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



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

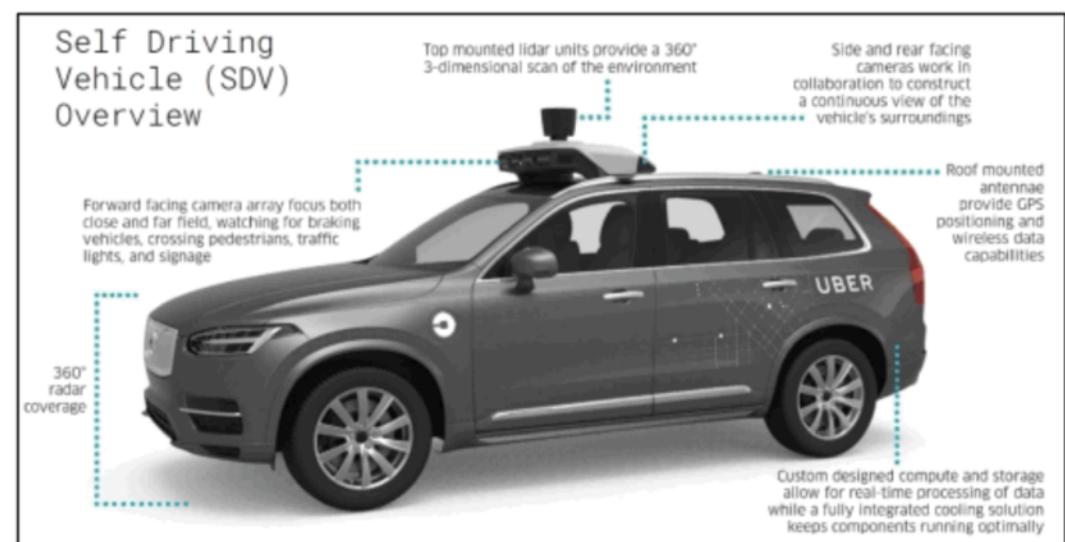
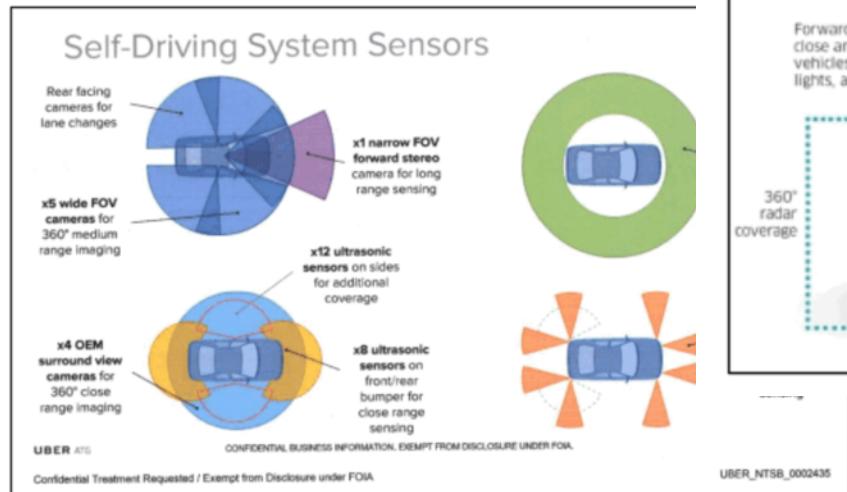
The system had several serious software flaws, the NTSB said.



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

25
Comments

1131
Shares



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.

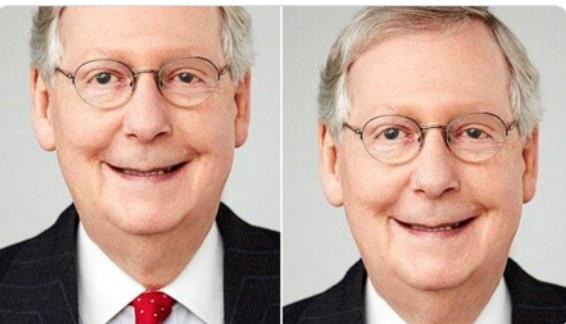
Sorry for the bias in the meanwhile. Have fun!

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**



Open Source Maintainers

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

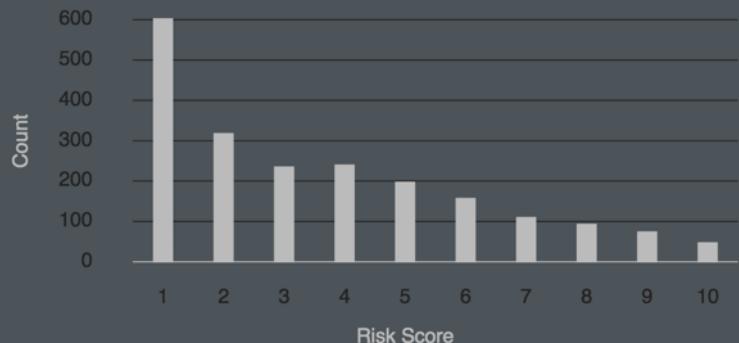
jaydenseric commented 6 days ago

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.

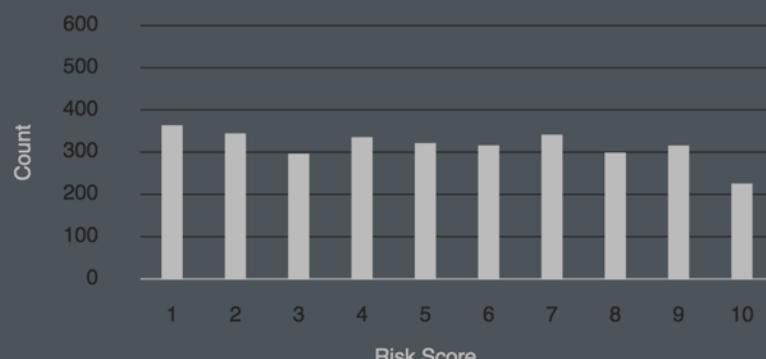
884 162 7 16 18

123abc

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

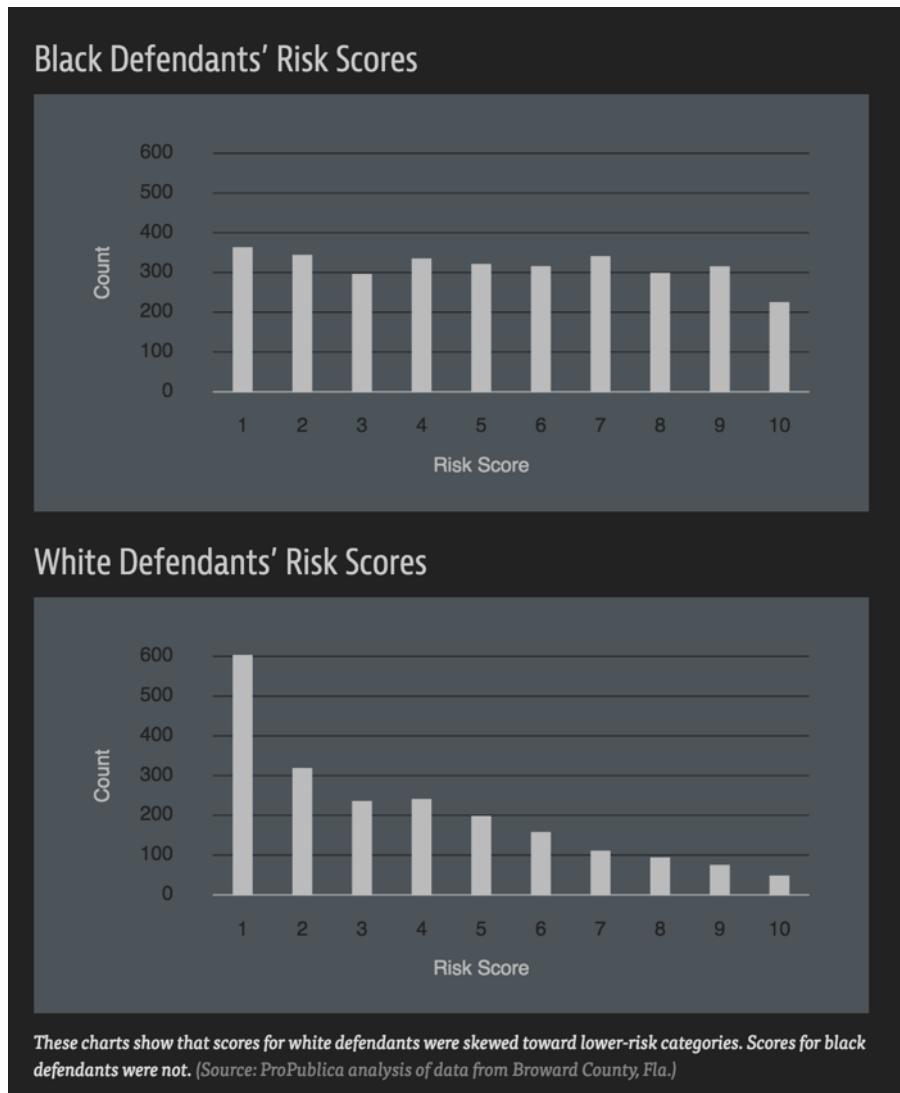
Access to money

Access to health care

Receiving parole

Possibility of Bail

Risk Scores



Therac-25

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

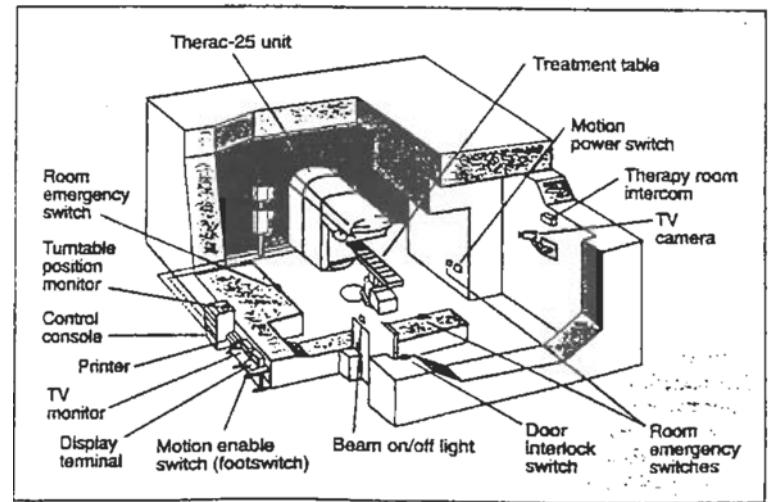
Traced to:

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	
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: 11:48:58	TREAT: TREAT PAUSE	X-RAY	173777
OPE ID: 033-tfs3p	REASON: OPERATOR	COMMAND: █	

The New York Times

Make a contribution

Subscribe Find

BUSINESS DAY

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

We Need To Work Harder To Make Software Engineering More Ethical

Jessica Baron Contributor @ Consumer Tech
I write about the ethics of science and technology.



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

US edition ▾

n

it ethics

READ

to fool AI with magic

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

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.

"We found that explicitly instructing participants to consider the ACM code of ethics in their decision making had no observed effect when compared with a control group."

Challenge:

How do we apply ethics to a field (Software Engineering) that 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)

<p>Use this worksheet to identify opportunities for Humane Technology.</p> <p>Product or feature: _____</p> <p>Value proposition: _____</p> <p>Measure of success: _____</p>			<p>What are Human Sensitivities?</p> <p><i>Human Sensitivities</i> are instincts that are often vulnerable to new technologies.</p>	
Human Sensitivity	We are inhibited when	What inhibits	We are supported when	Opportunity to improve
Emotional 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"> Artificial scarcity Urgency signalling Constant monitoring Optimizing for screentime 	Design engenders calm, balance, safety, pauses and supports circadian rhythms.	<input type="radio"/> High <input type="radio"/> Low
Attention How and where we focus our attention.	Attention is physiologically drawn, overwhelmed or fragmented.	<ul style="list-style-type: none"> Constant context switching Many undifferentiated choices Fearful information No stopping cues (e.g. infinite scroll) Unnecessary movement 	Enabled to bring more focus and mindfulness.	<input type="radio"/> <input type="radio"/>
Sensemaking 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"> Facts out of context Over-personalized filters Equating virality with credibility Deceptive authority (ads vs. content) 	Enabled to consider, learn, express and feel grounded.	<input type="radio"/> <input type="radio"/>
Decisionmaking How we align our actions with our intentions.	Intentions and agency are not solicited nor supported.	<ul style="list-style-type: none"> Avatars to convey authority Stalking ads and messages Push content models Serving preference over intent 	Enabled to gain agency, purpose, and mobilization of intent.	<input type="radio"/> <input type="radio"/>
Social Reasoning How we understand and navigate our personal relationships.	Status, relationships or self-image are manipulated.	<ul style="list-style-type: none"> Quantified social status Viral sharing Implied obligation Enabling impersonation 	Enabled to connect more safely and authentically with others.	<input type="radio"/> <input type="radio"/>
Group Dynamics How we navigate larger groups, status, and shared understanding.	Excluded, divided or mobilized through fear.	<ul style="list-style-type: none"> Suppressing views and nuance Enabling ad hominem or hate speech Enabling viral outrage Lack of agreed-upon norms 	Enabled to develop a sense of belonging and cooperation.	<input type="radio"/> <input type="radio"/>

[Center for Humane Technology] 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
Attention How and where we focus our attention.	Attention is physiologically drawn, overwhelmed or fragmented.	<ul style="list-style-type: none">• Constant context switching• Many undifferentiated choices• Fearful information• No stopping cues (e.g. infinite scroll)• Unnecessary movement	Enabled to bring more focus and mindfulness.	

Identify Opportunities to improve

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 Jones¹



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 move 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.

¹ 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

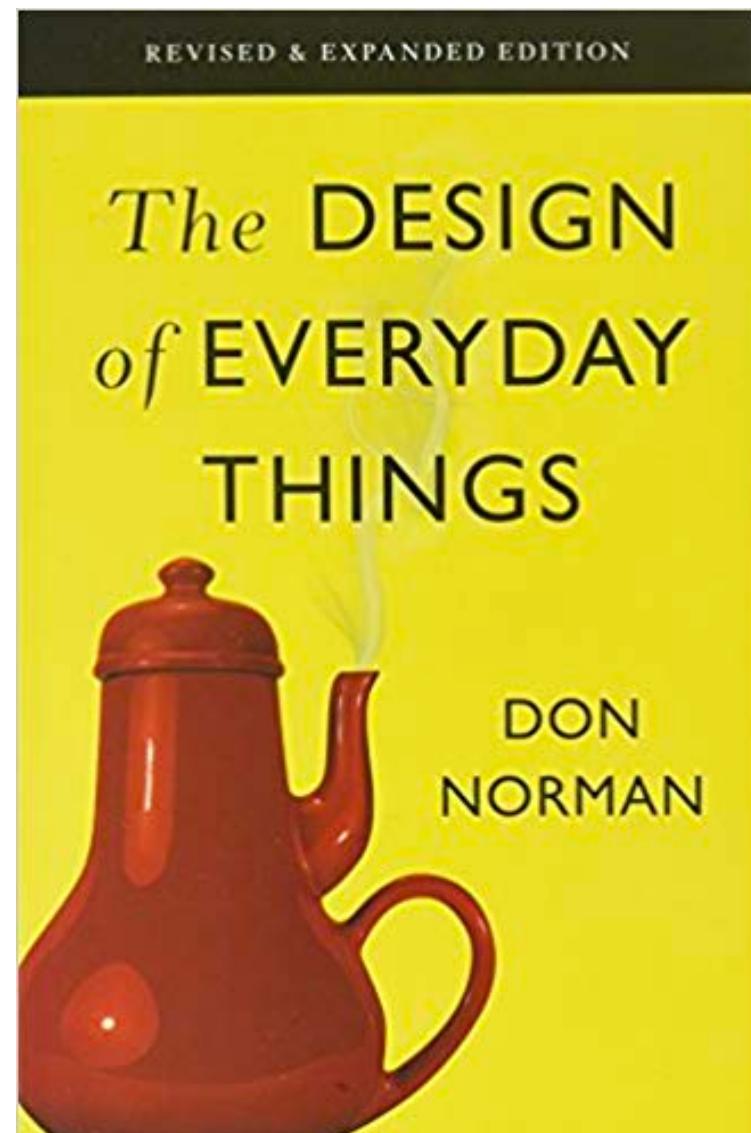
<https://gendermag.org>

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

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



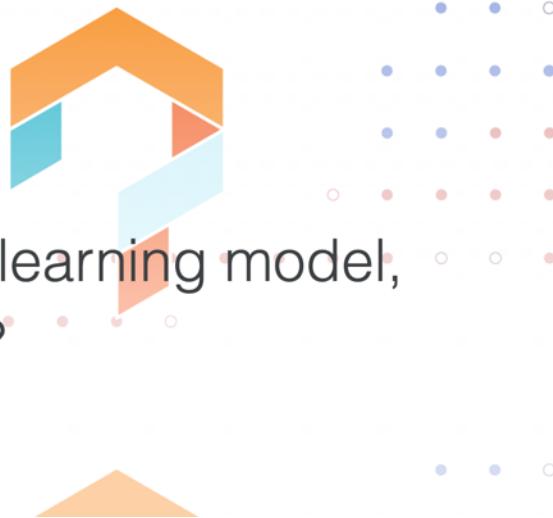
**2.Does my
software amplify
positive or
negative behavior
for users and
society at large?**

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 Co... Binning | Y-Axis C... Color By
age 10 marital-stat 1 Inference

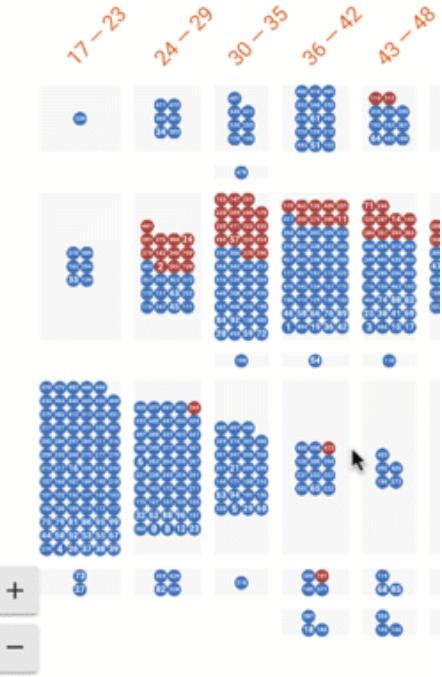
Select a datapoint to begin exploring features and values. →

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.

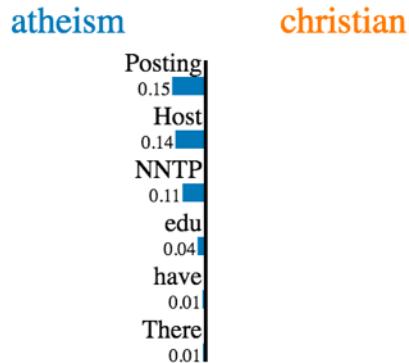
Use the Features tab to view statistics about your dataset.



Local Interpretable Model-Agnostic Explanations LIME)

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

Prediction probabilities



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

Why you're seeing this ad

 Only you can see this

Simplifi Money wants to reach people like you, who may have:

-  Set their age between 25 and 50 >
-  A primary location in the United States >

What else influences your ads

Your ads may be based on other advertiser choices, your profile and activities—like websites you visit and ads you interact with—as well as other information not listed here. [Learn more about how ads work](#)

What you can do

-  Hide all ads from this advertiser Hide
You won't see Simplifi Money's ads
-  Make changes to your ad preferences >
Adjust settings to personalize your ads

Learn about your privacy at Meta

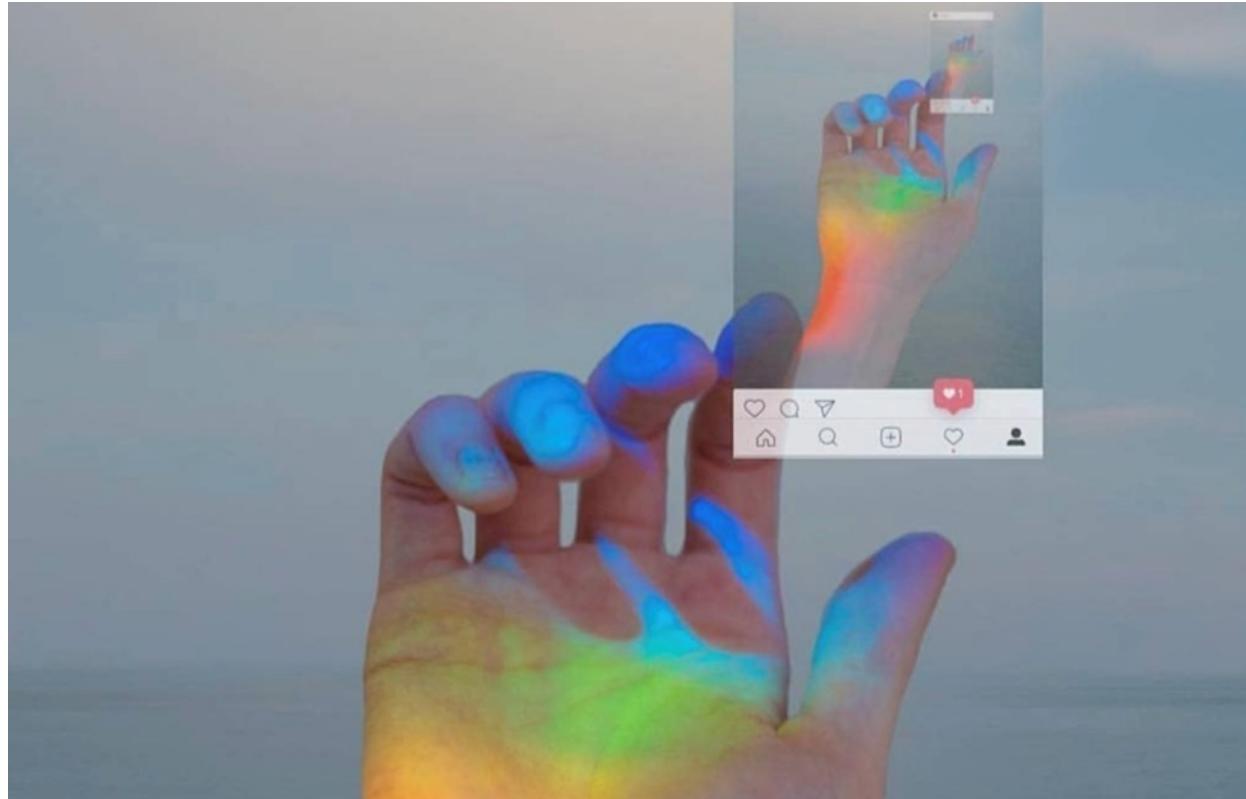
We want to help you understand how Meta uses your information to show you ads.

How businesses use our ads system

Businesses try to reach people based on their interests, characteristics, or using information about who visits the business's website...

You have options to manage the ads you see on Facebook

To give you more control of the ads you see, we have a number of tools to manage your ad experience.



@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/

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>



| 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

The screenshot shows a web browser window for the "Welcome to the Artificial Intelligence Incident Database". The URL in the address bar is "incidentdatabase.ai". The page features a large "AI" logo and the text "INCIDENT DATABASE". On the left sidebar, there are links for "Discover" and "Submit", followed by a list of items under "Database Apps": "Discover App", "Incident Report Submission", and "Your App Here". Below this are links for "About Us" and "Contact and Follow", and a link to "Partnership on AI Home". The main content area has a heading "Welcome to the AIID" and a section titled "Why 'AI Incidents'?". It discusses the potential dangers of intelligent systems and the need for a repository of incidents. Another section, "What is an Incident?", is partially visible below it. A sidebar on the right is titled "CONTENTS" and lists several topics: "Why 'AI Incidents'?", "What is an Incident?", "Current and Future Users", and "When Should You Report an Incident?".

Welcome to the AIID

Why "AI Incidents"?

Intelligent systems are currently prone to unforeseen and often dangerous failures when they are deployed to the real world. Much like the transportation sector before it (e.g., [FAA](#) and [FARS](#)) and more recently [computer systems](#), intelligent systems require a repository of problems experienced in the real world so that future researchers and developers may mitigate or avoid repeated bad outcomes.

What is an Incident?

The initial set of more than 1,000 incident reports have been intentionally broad in nature. Current examples include,

**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
- composable [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
- manageable
- 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
- tailorable
- 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

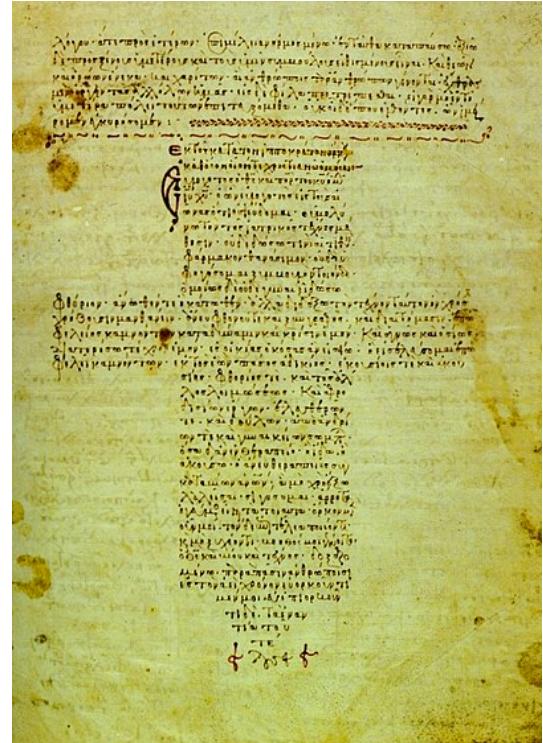


Medicine - Intrinsic

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

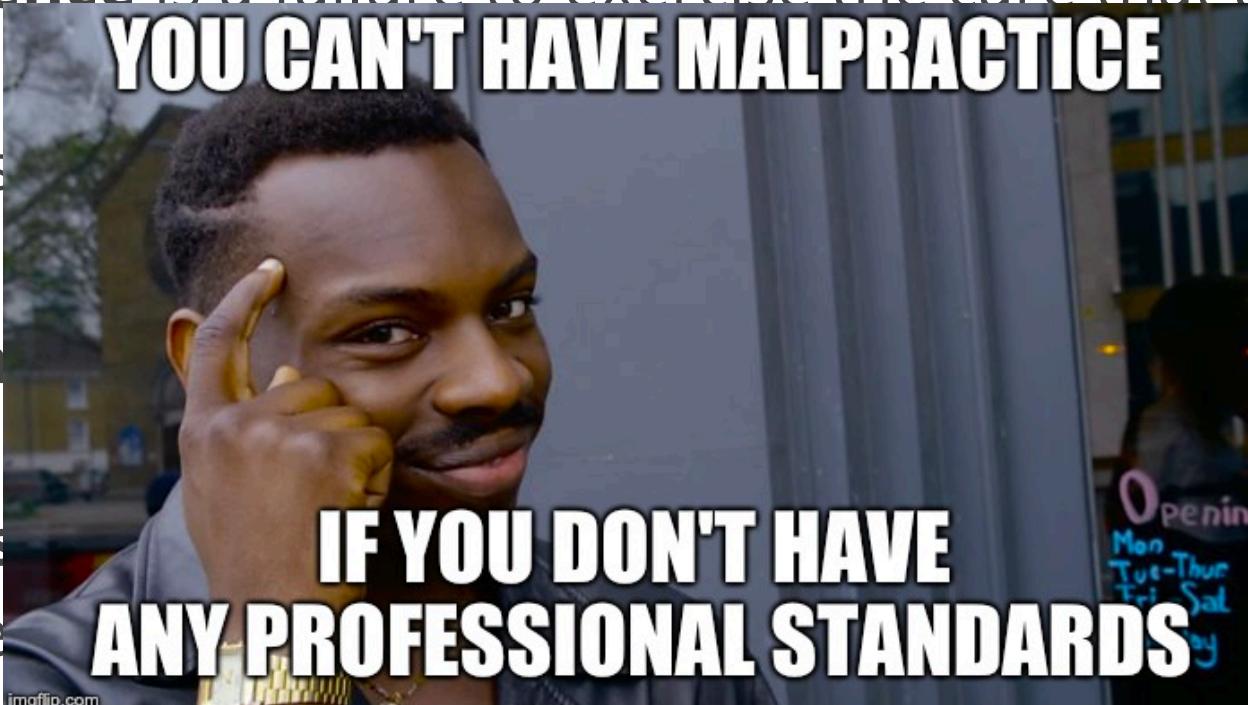
due to negligence or with intent to harm

And cause damages to a client

Malpractice vs. Negligence

Negligence is a failure to exercise the care that a reasonable person would exercise under the same circumstances.

Malpractice is negligence by a "professional" (e.g., a professional who provides services to the public) that fails to meet the standards of practice set by a governing body ("standard of care"), subsequently causing harm to the plaintiff.

A meme featuring a man from the TV show 'Key & Peele'. He is smiling and holding his chin with his hand. Overlaid on the image is the text 'YOU CAN'T HAVE MALPRACTICE' in large, bold, white letters. Below this, in smaller white letters, is 'IF YOU DON'T HAVE ANY PROFESSIONAL STANDARDS'. The image is set against a background of a city street with buildings and a sign that says 'Opening Mon Tue-Thur Fri-Sat Sun'.

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>

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?