

Ethics

17-313 Spring 2025

Foundations of Software Engineering

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

Michael Hilton, Austin Henley, and Nadia Nahar

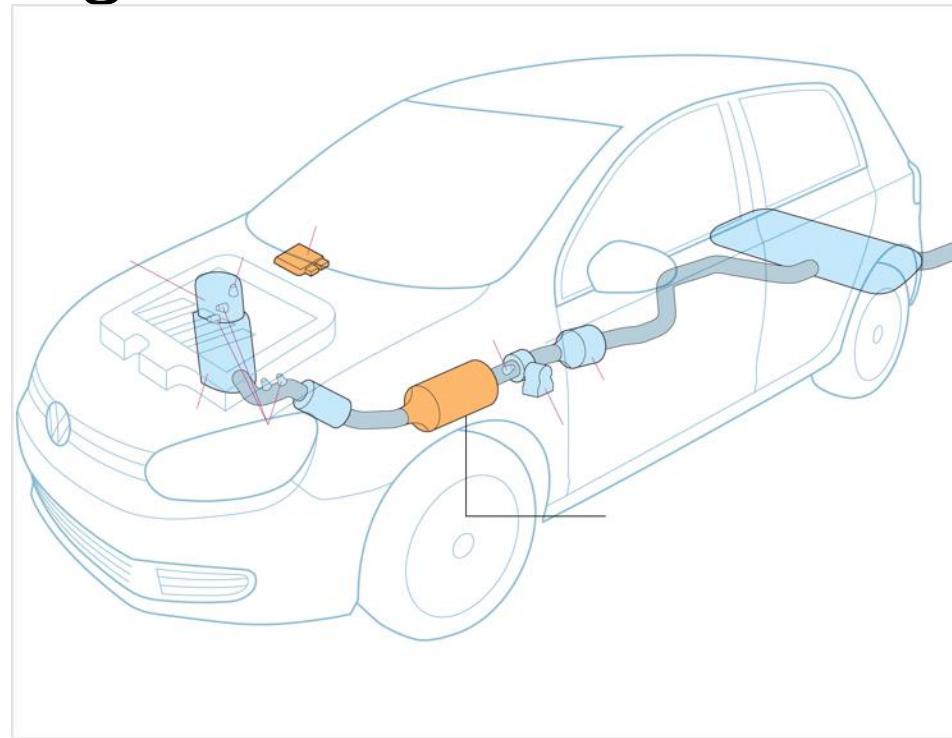
Administrivia

- Final Project Presentation during our final exam time.
 - Monday, May 5, 2025 01:00pm - 04:00pm

Ethics

Volkswagen Scandal

VW was caught cheating on emissions for Diesel engines



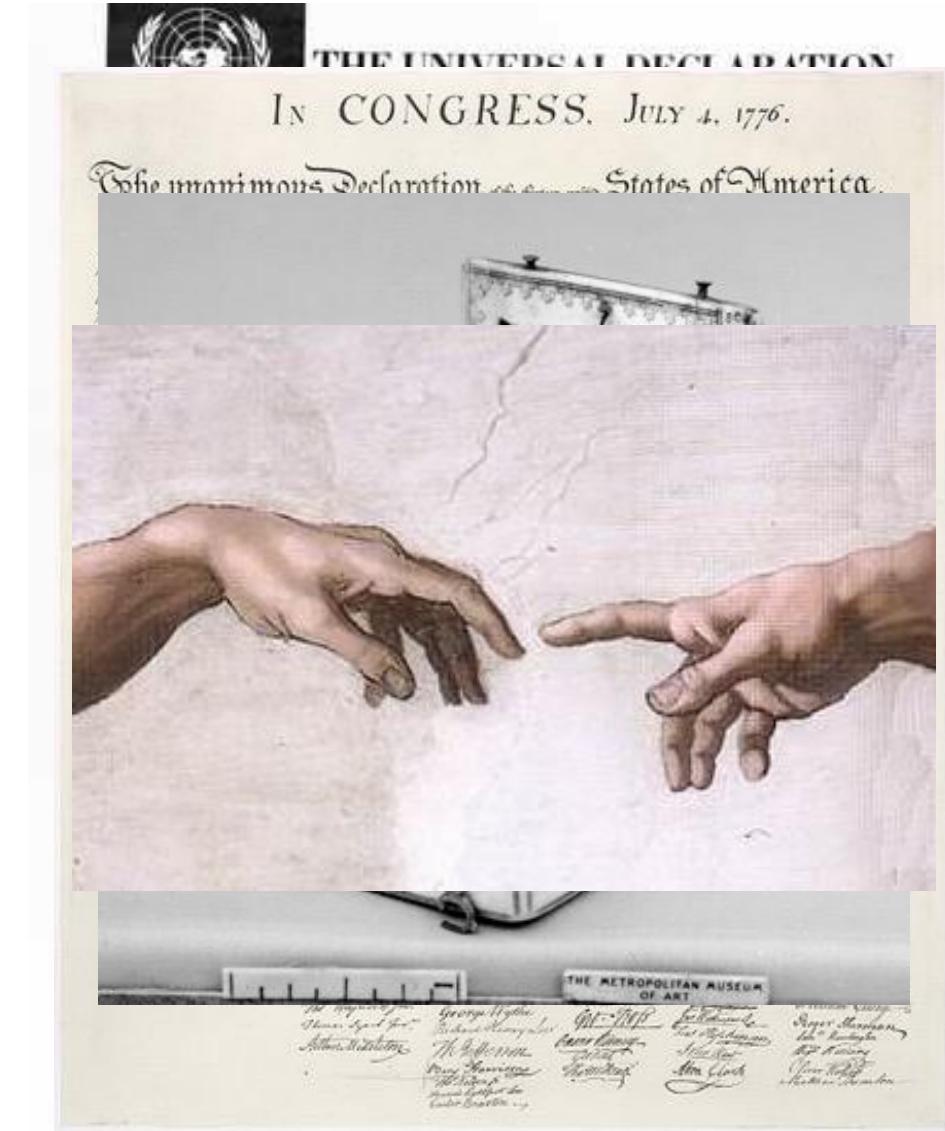
<https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html?mtrref=www.google.com&assetType=REGIWALL>

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



UNITED NATIONS

Activity: (Un)Ethical situations

80

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

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 | Nov 27, 2018, 6:10pm EST

f t  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

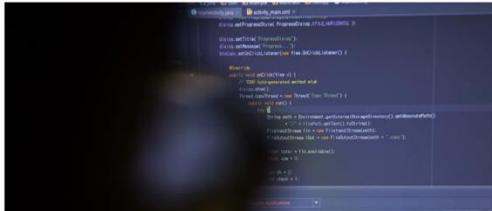


Open Intellectual Property Concerns

- Was the data used to train these LLMs obtained illegally?
- Who owns the IP associated with LLM outputs?
- Should sensitive information be provided as inputs to LLMs?

ARTIFICIAL INTELLIGENCE / TECH / LAW

The lawsuit that could rewrite the rules of AI copyright



/ Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI. But the suit could have a huge impact on the wider world of artificial intelligence.

ARTIFICIAL INTELLIGENCE / TECH / CREATORS

AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit



/ The suit claims generative AI art tools violate copyright law by scraping artists' work from the web without their consent.

By James Vincent, a senior reporter who has covered AI robotics, and more for eight years at The Verge.
Jan 16, 2023, 6:28 AM EST | □ 27 Comments / 27 New



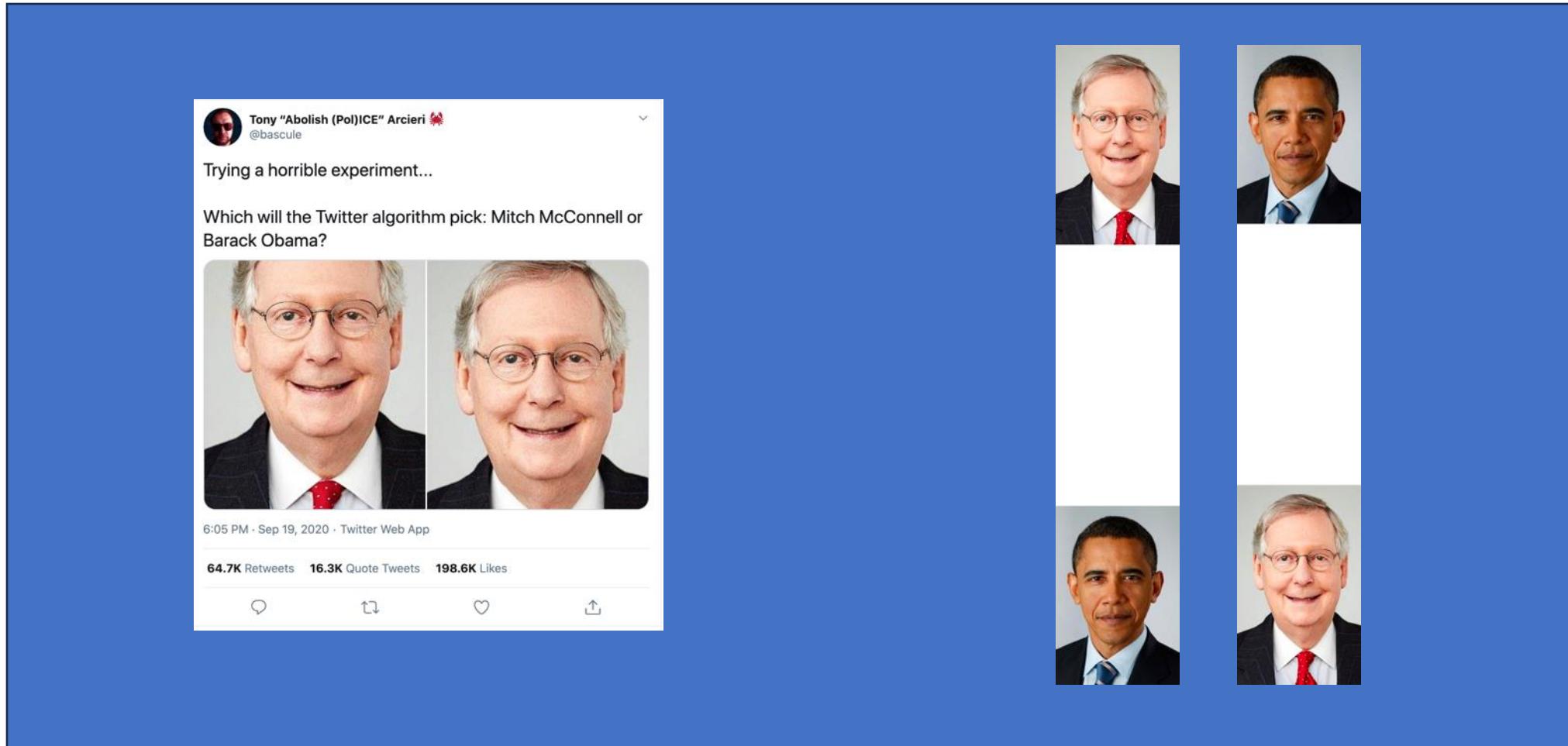
Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT

ChatGPT doesn't keep secrets.

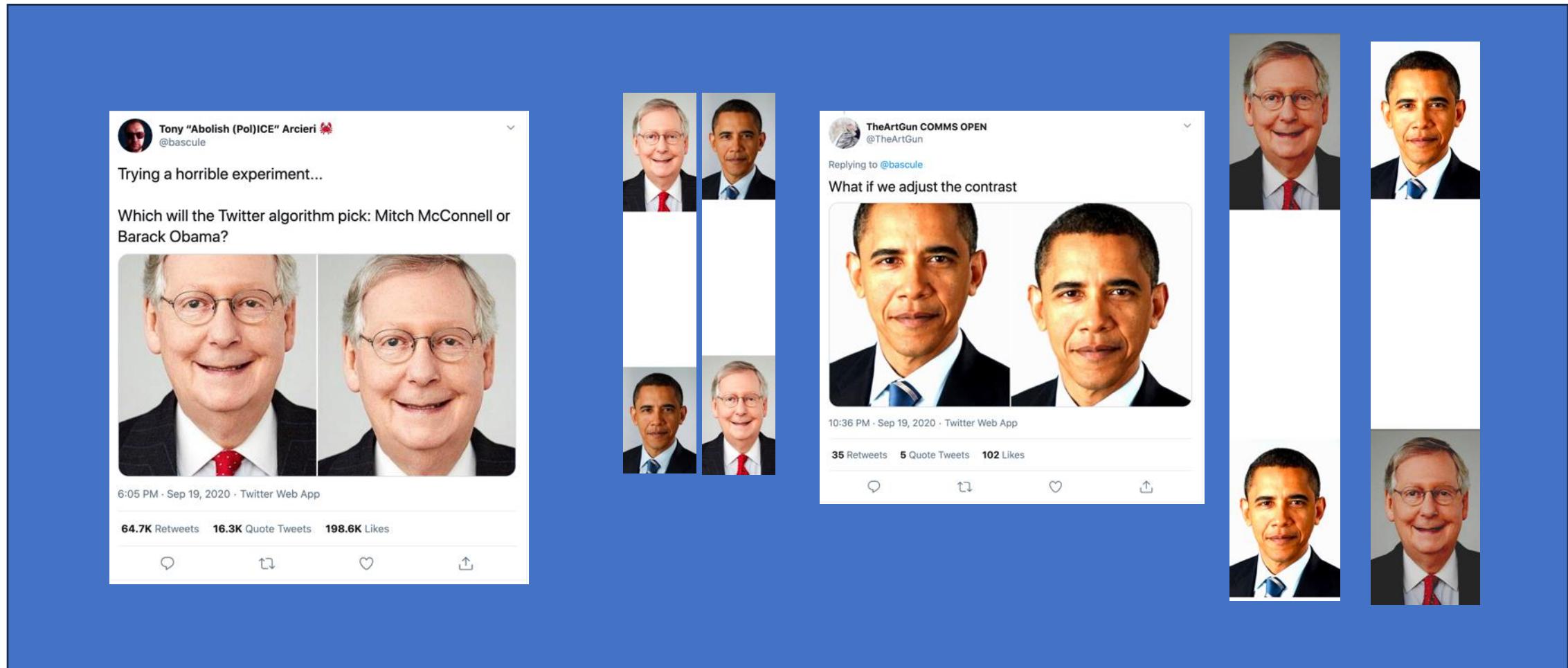
By Cecily Mauran on April 6, 2023



Twitter cropping photos



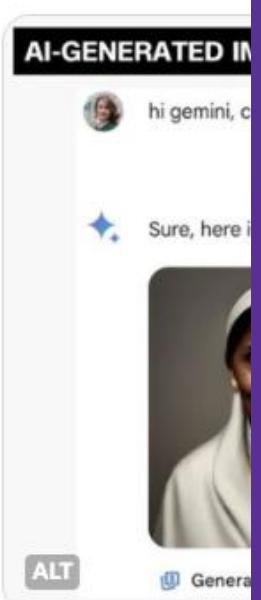
Twitter cropping photos



MMitchell

@mmitchell_a

I really love the act
Google Gemini's te
representation. As
world (>4 years! ha



4:08 PM · Feb 25, 2024

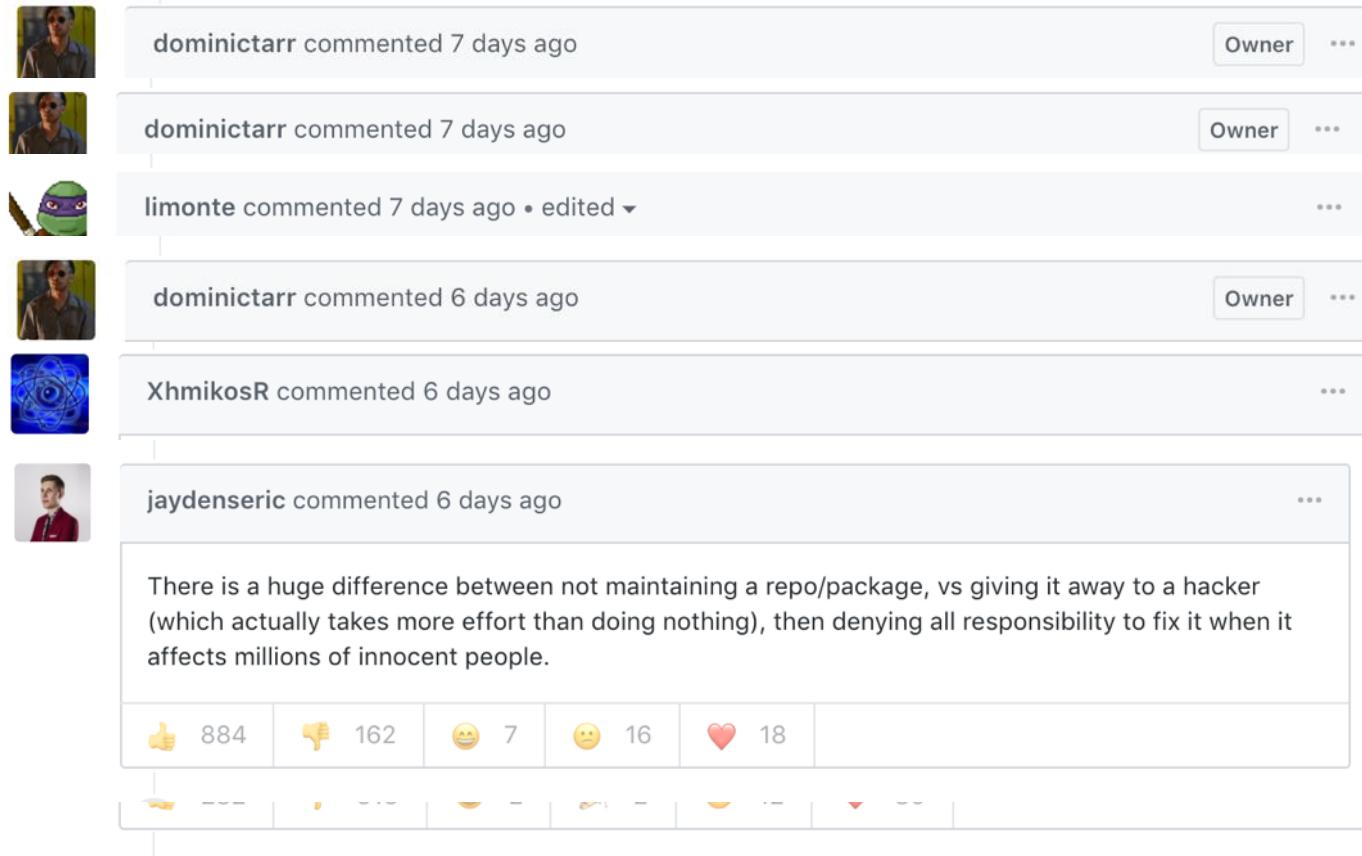
Foresight in AI Chart: Text-to-Image example

Some (of many) example people and use contexts		People			
		Users		Those affected	
Use Contexts	Intended	Intended	Unintended	Intended	Unintended
	Historic bias depiction, Dream world depiction	😊		😊	
	Mis/disinformation (e.g., Deep Fakes)		Revenge porn		Creative content that should be paid for is plagiarised
Out of scope		System won't work			

CC BY / Margaret Mitchell / m-mitchell.com

36

Open Source Maintainers



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

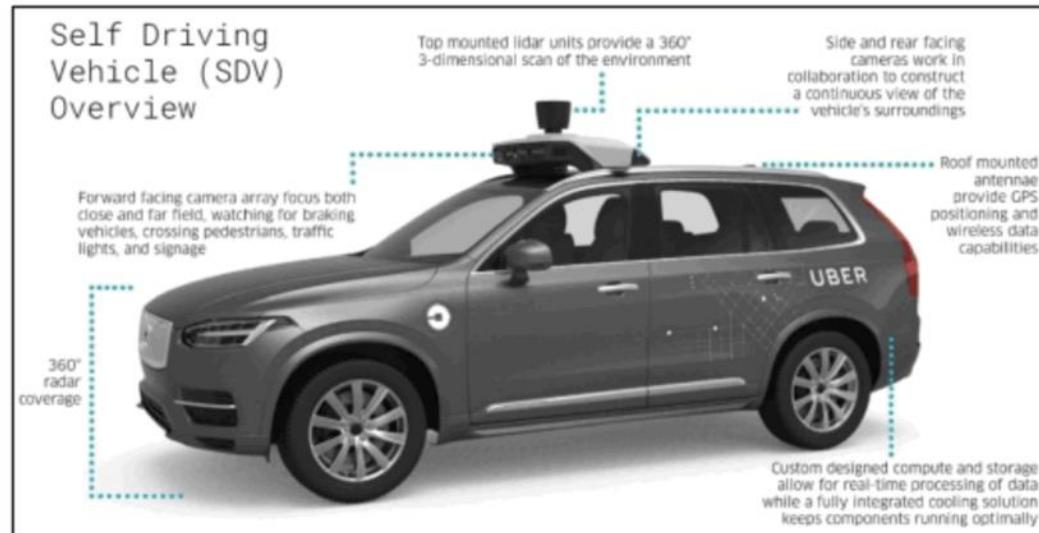
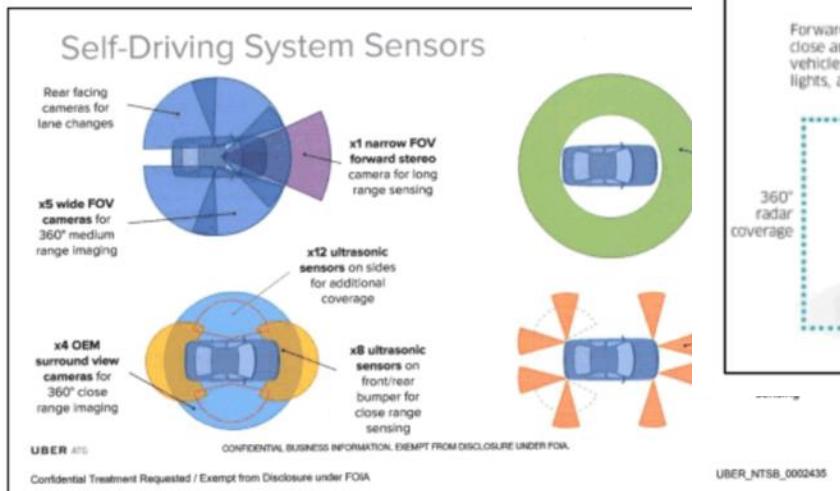
The system had several serious software flaws, the NTSB said.



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

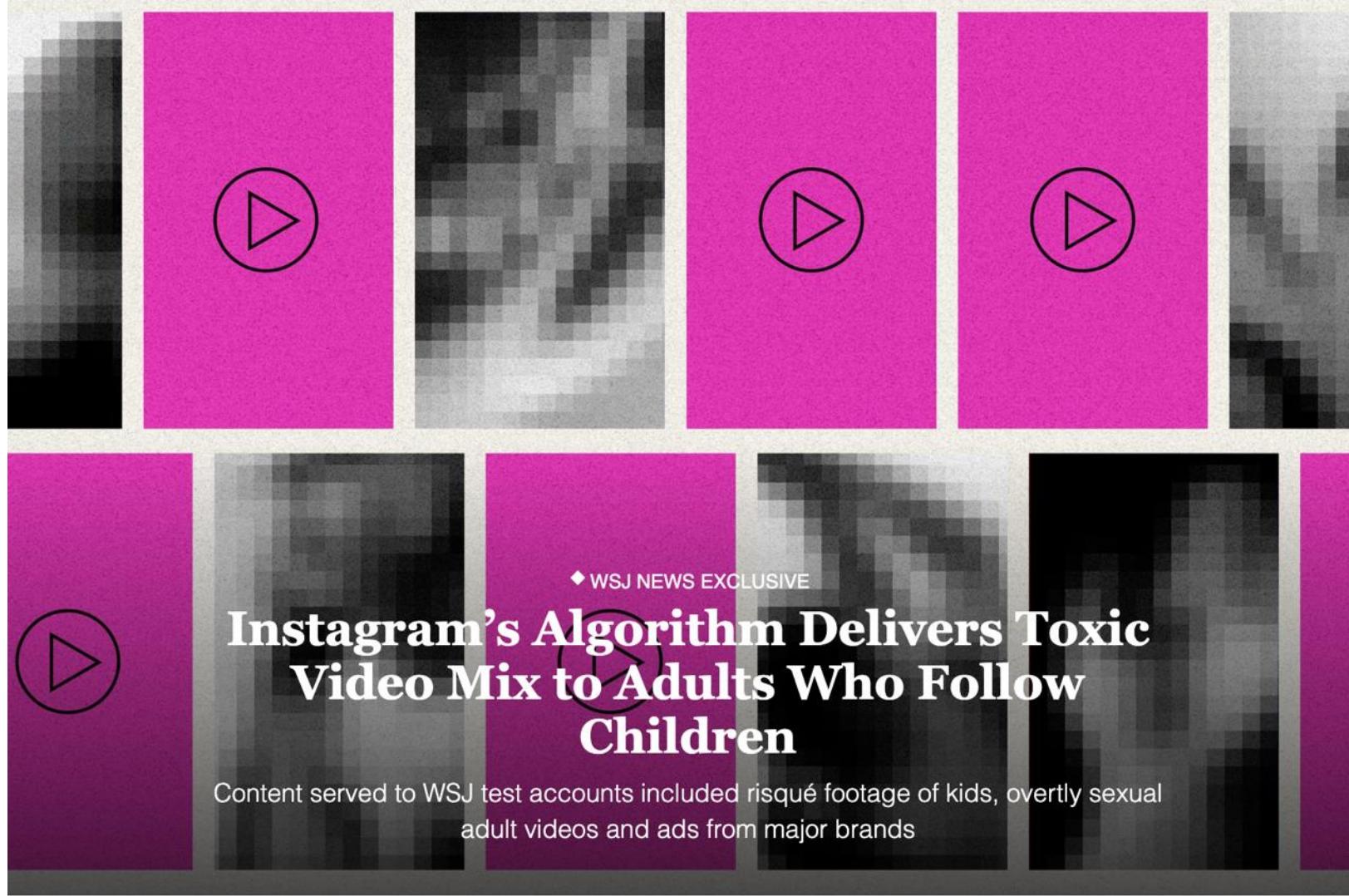
25
Comments

1131
Shares





THE WALL STREET JOURNAL.



DAISY KORPICS FOR THE WALL STREET JOURNAL





tomorrow
belongs to those who embrace it
today

[trending](#)[tech](#)[innovation](#)[business](#)[security](#)[advice](#)[buying guides](#)

/ innovation

[Home](#) / [Innovation](#) / [Artificial Intelligence](#)

Microsoft adds new Designer protections following Taylor Swift deepfake debacle

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

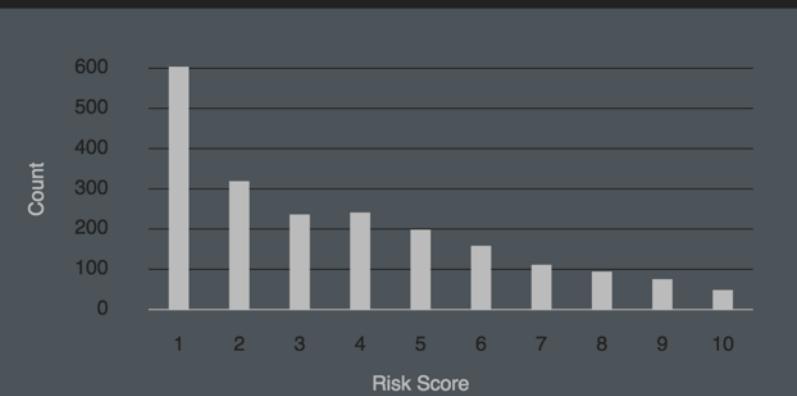


Written by **Sabrina Ortiz**, Editor

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

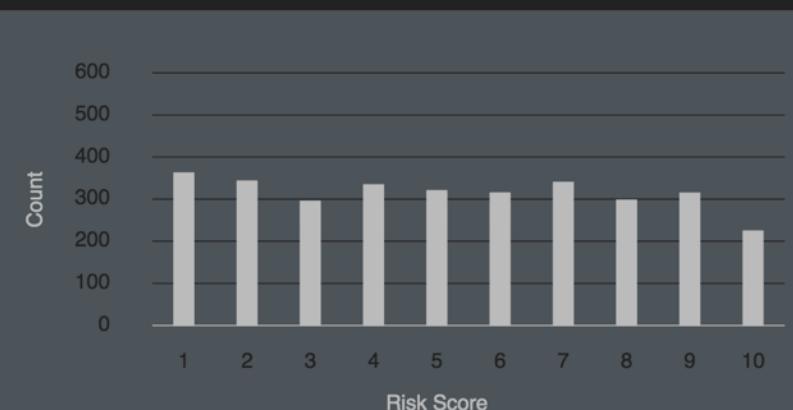


White Defendants' Risk Scores



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

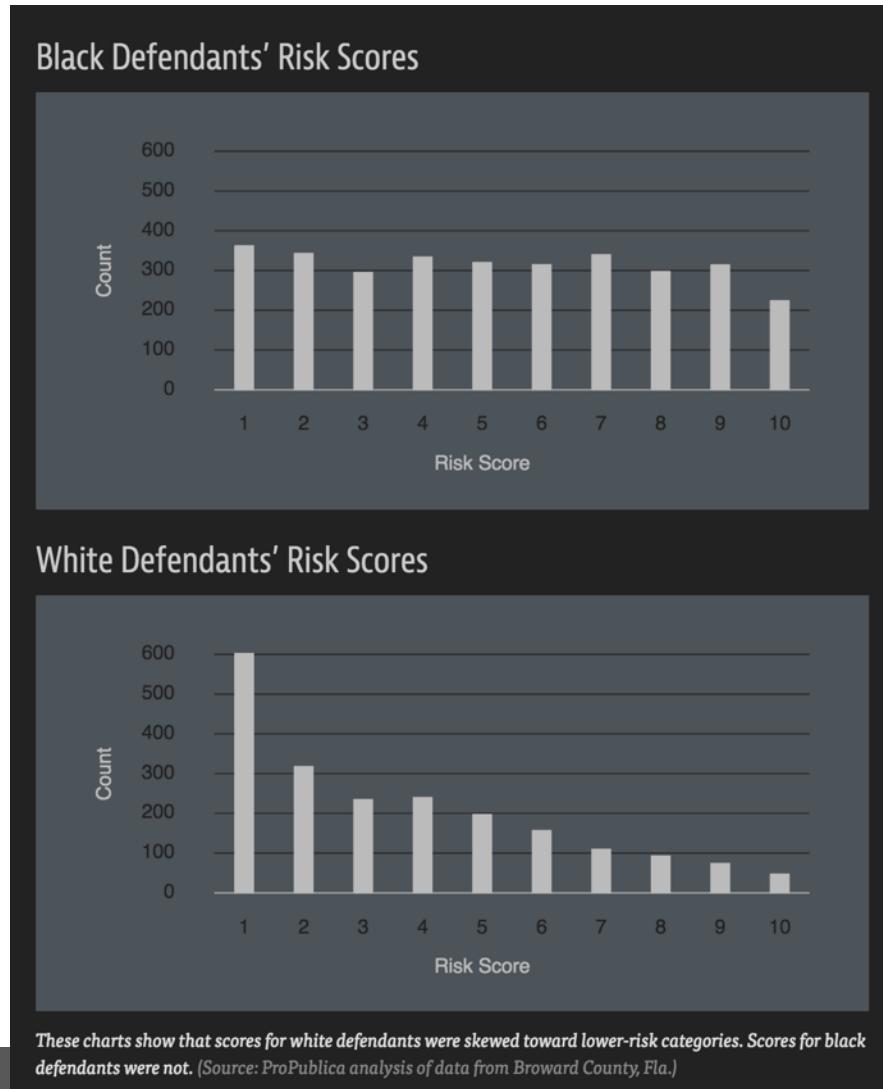
Access to money

Access to health care

Receiving parole

Possibility of Bail

Risk Scores



Therac-25

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

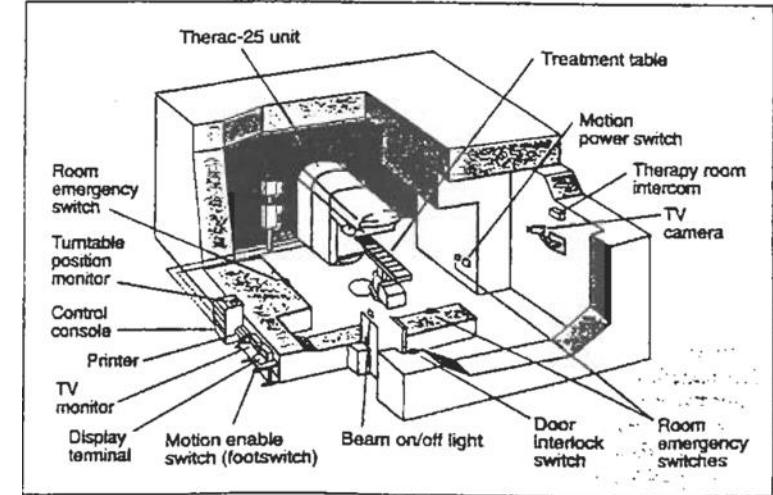
Traced to:

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
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
OFR ID: 033-tfe3p	REASON: OPERATOR	COMMAND:	[redacted]

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

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 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.		What are Human Sensitivities?	
Product or feature:		Human Sensitivity	We are inhibited when
Value proposition:			What inhibits
Measure of success:			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 scarcityUrgency signallingConstant monitoringOptimizing for screen time
		Attention How and where we focus our attention.	Attention is physiologically drawn, overwhelmed or fragmented. <ul style="list-style-type: none">Constant context switchingMany undifferentiated choicesFearful informationNo stopping cues (e.g. infinite scroll)Unnecessary movement
		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 contextOver-personalized filtersEquating virality with credibilityDeceptive authority (ads vs. content)
		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 authorityStalking ads and messagesPush content modelsServing preference over intent
		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 statusViral sharingImplied obligationEnabling impersonation
		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 nuanceEnabling ad hominem or hate speechEnabling viral outrageLack of agreed-upon norms

[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

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 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:

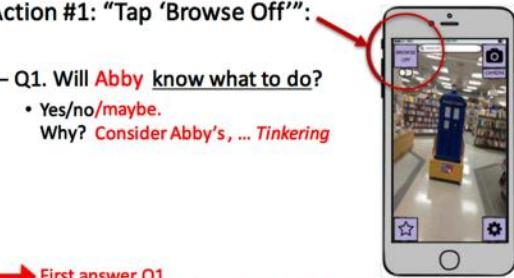
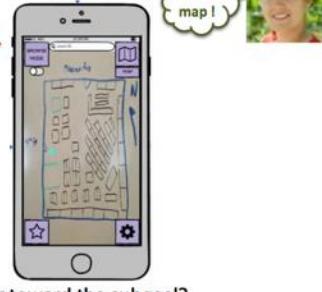
- **Information Processing Style:** Abby tends towards a comprehensive information processing style when she needs to move information. So, instead of acting upon the first option that seems promising, she gathers information comprehensively to try to form a complete understanding of the problem before 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.

¹Abby 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

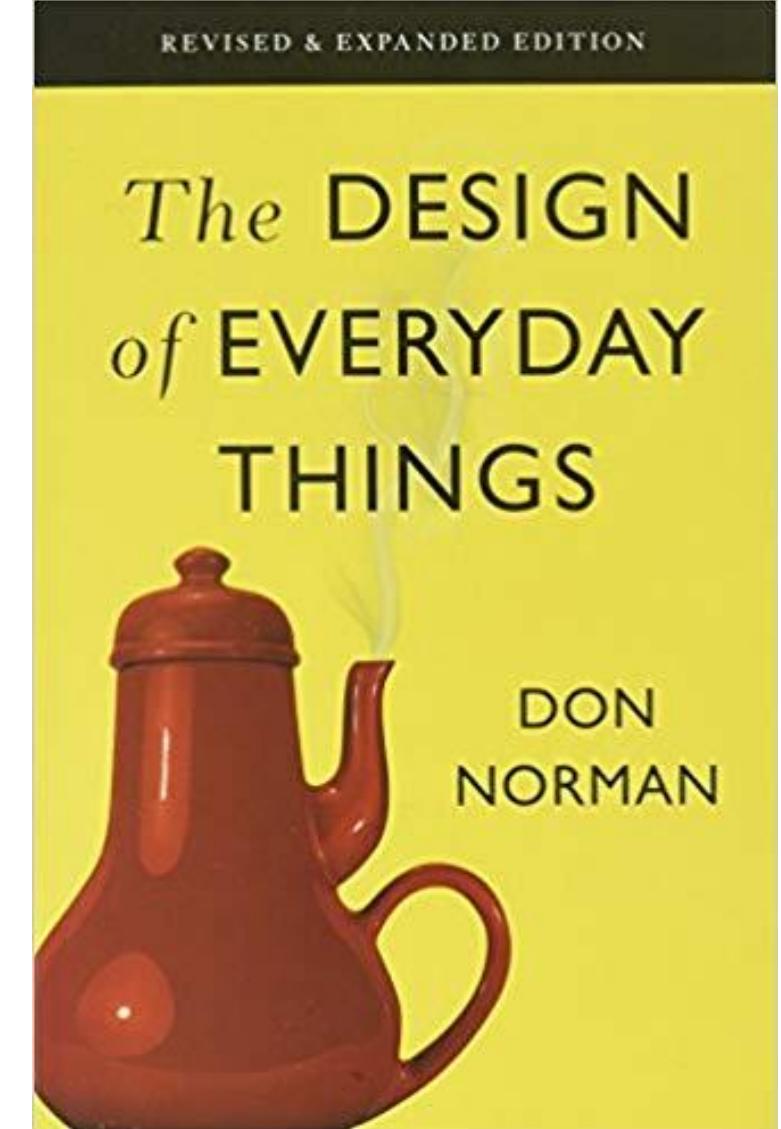
<https://gendermag.org>

<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...?<ul style="list-style-type: none">• 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?<ul style="list-style-type: none">• 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?<ul style="list-style-type: none">• 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

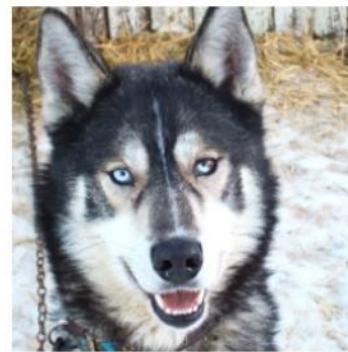


Agile

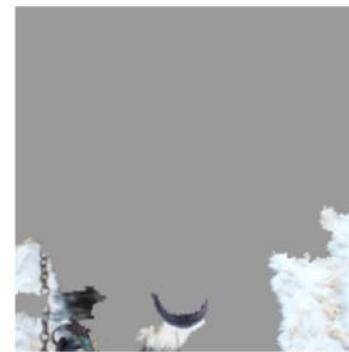


2.Does my software amplify
positive or negative behavior for
users and society at large?

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)

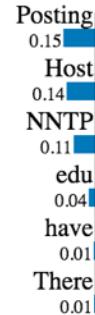
<https://github.com/marcotcr/lime>

Prediction probabilities

atheism	0.58
christian	0.42

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

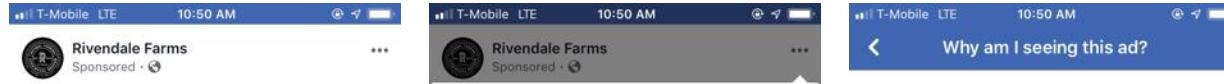
NNTP-Posting-Host: triton.unm.edu

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



There may be other reasons you're seeing this ad, including that Rivendale Farms wants to reach **people ages 22 to 64 who live or were recently near Pittsburgh, Pennsylvania**. This is information based on your Facebook profile and where you've connected to the internet.





@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>



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

The screenshot shows a web browser window for the AI Incident Database (AIID) at incidentdatabase.ai. The page has a dark blue header with the AIID logo and navigation links for 'Discover' and 'Submit'. The main content area features a 'Welcome to the AIID' section with a large heading and a 'Why "AI Incidents"' section with explanatory text. A sidebar on the right lists 'CONTENTS' including 'Why "AI Incidents"?' and 'What is an Incident?'. The footer contains links for 'Discover App', 'Incident Report Submission', and 'Your App Here'.

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
- 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.

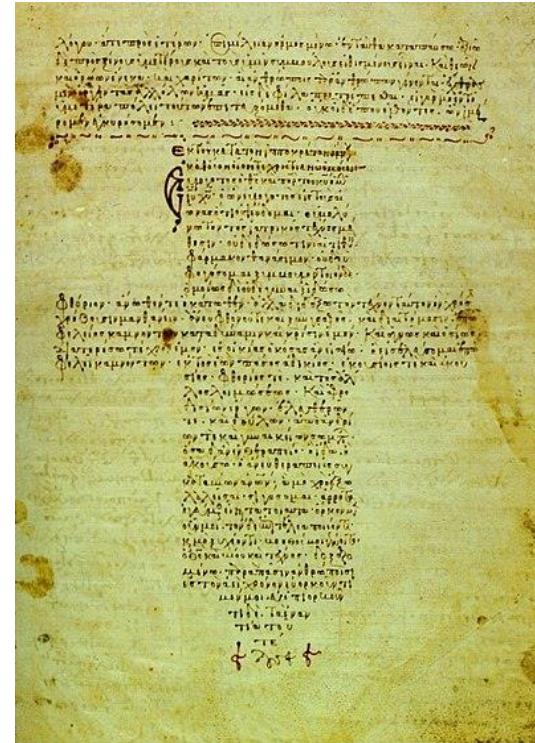
First three “professions”

- Divinity,
- Law
- Medicine



Medicine - Intrinsic

Hippocratic Oath ~450BC “Do no Harm”



Law -Extrinsic

Bar regulates behavior

Oath to follow rules

Malpractice



Legal Malpractice

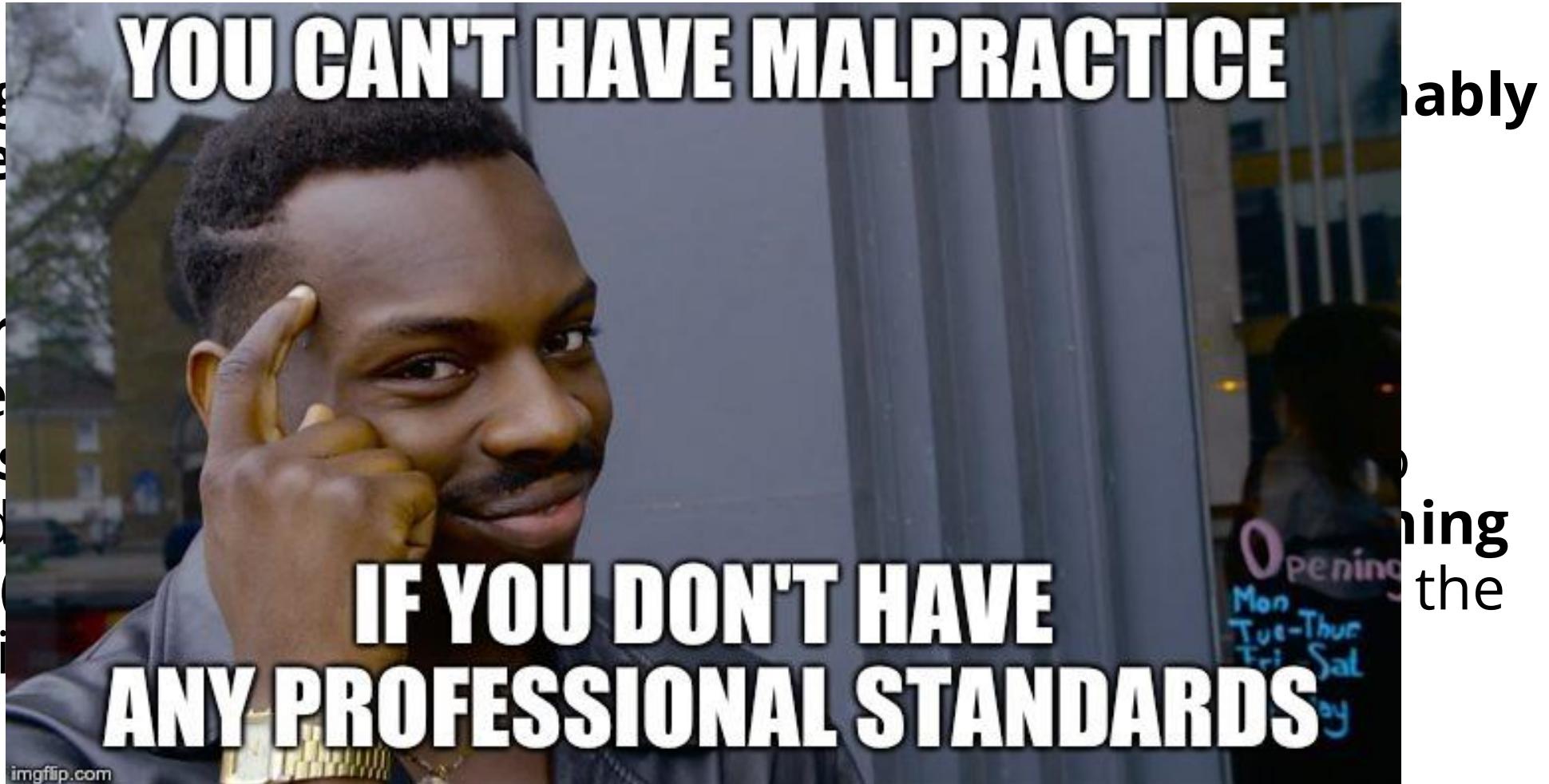
Not every mistake is legal malpractice. For malpractice to exist:

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

Malpractice vs. Negligence

Negligent
prudent

Malpractice
"professional
professionals
provided
body of
plaintiffs



imgflip.com

DISCUSSION: What should we do going forward?

Bioengineering Ethics:

- Respect for Autonomy
- Beneficence
- Nonmaleficence
- Justice

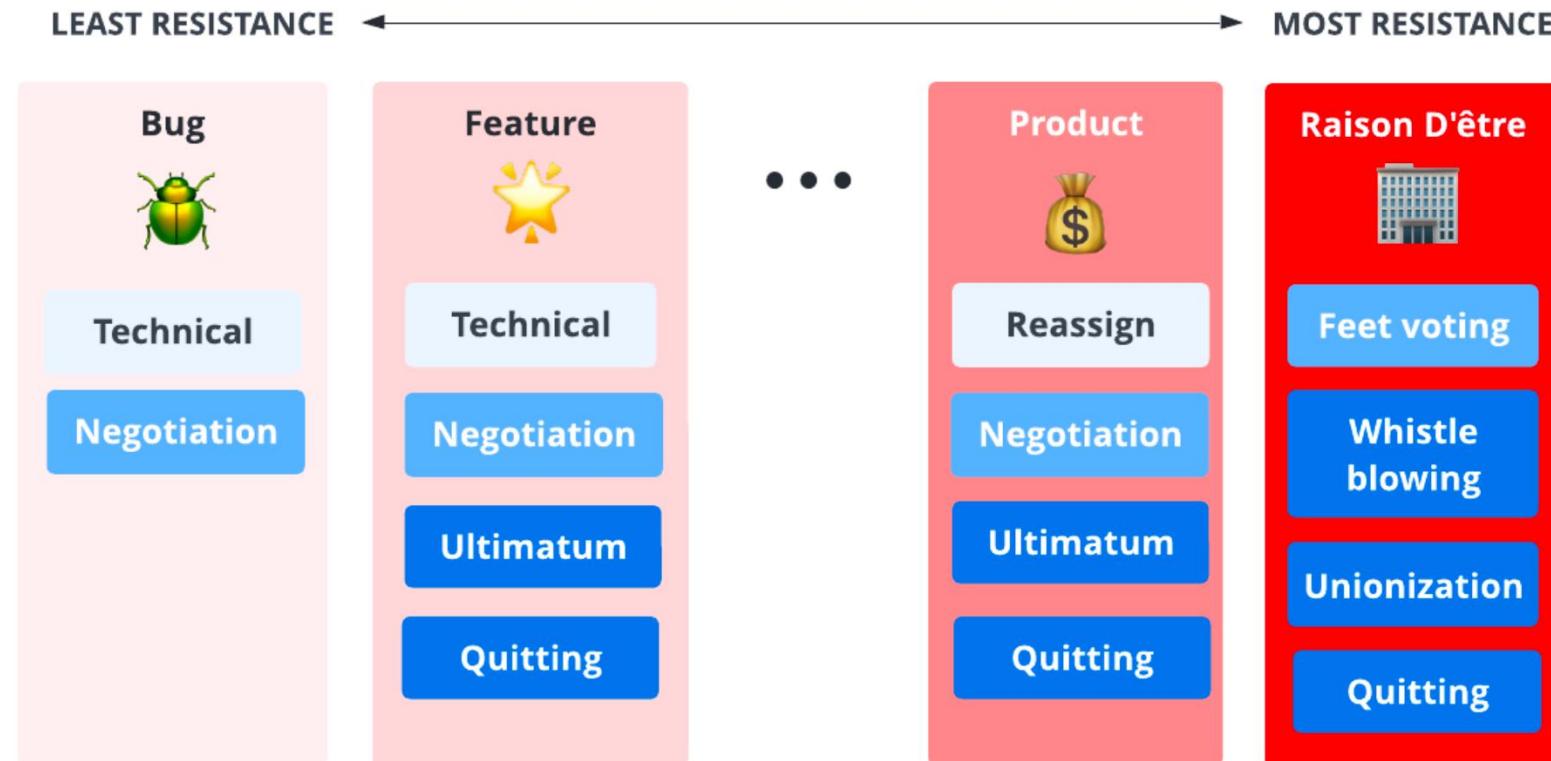
Professional Engineers

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?