Ethics

17-313 Spring 2024

Foundations of Software Engineering

https://cmu-313.github.io

Michael Hilton and Eduardo Feo Flushing

Administrivia

- Midterm is not graded yet. But will be included in mid-semester grade. P2C will not be included in midsemester grades.

- We are not expecting work during spring break. P3 will be assigned after spring break

Ethics

Volkswagen Scandal

VW was caught cheating on emissions for Diesel engines

S3D Software and Societal Systems Department

Carnegie Mellon University
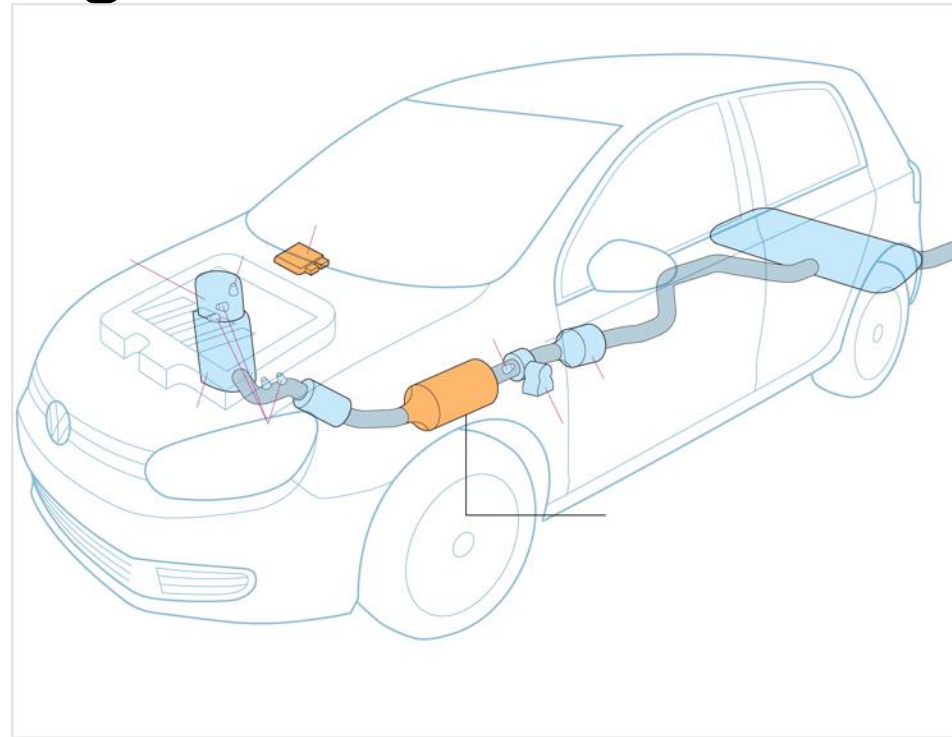
What is Human Flourishing?

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

'People like surprises,' executive tells UK Parliament

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

f 🐦 ↗ SHARE

80 💬

Software and Societal
Systems Department

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

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

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

f 🐦 ↗ SHARE

S3D Software and Societal Systems Department

Carnegie Mellon University

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

S3D Software and Societal Systems Department

Carnegie Mellon University

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

S3D Software and Societal Systems Department

Carnegie Mellon University

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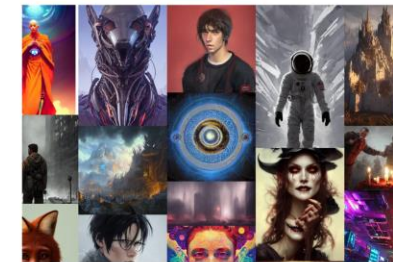
/ The suit claims generative AI art tools violate copyright law by scraping artists' work from the web without their consent.

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

A collage of AI-generated images created using Stable Diffusion. Image: The Verge via Lexica

Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT
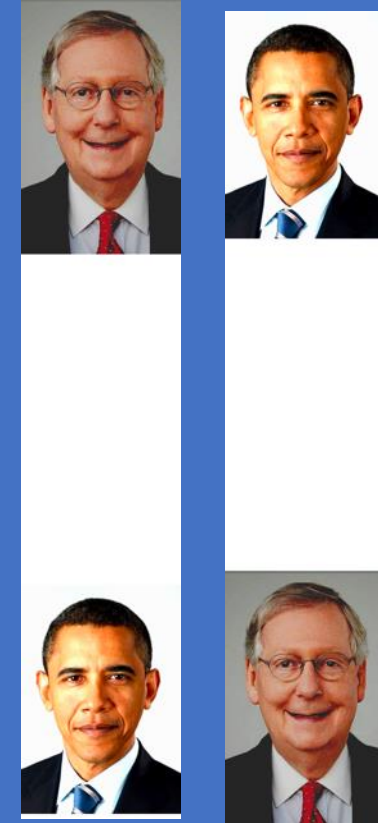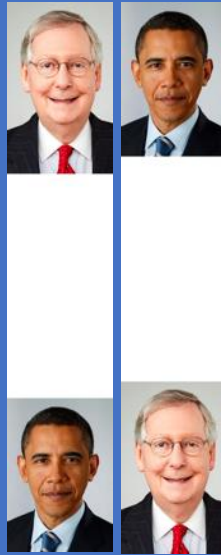
ChatGPT doesn't keep secrets.

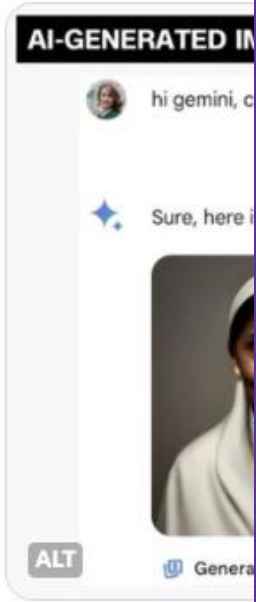By Cecily Mauran on April 6, 2023

Twitter cropping photos

Twitter cropping photos
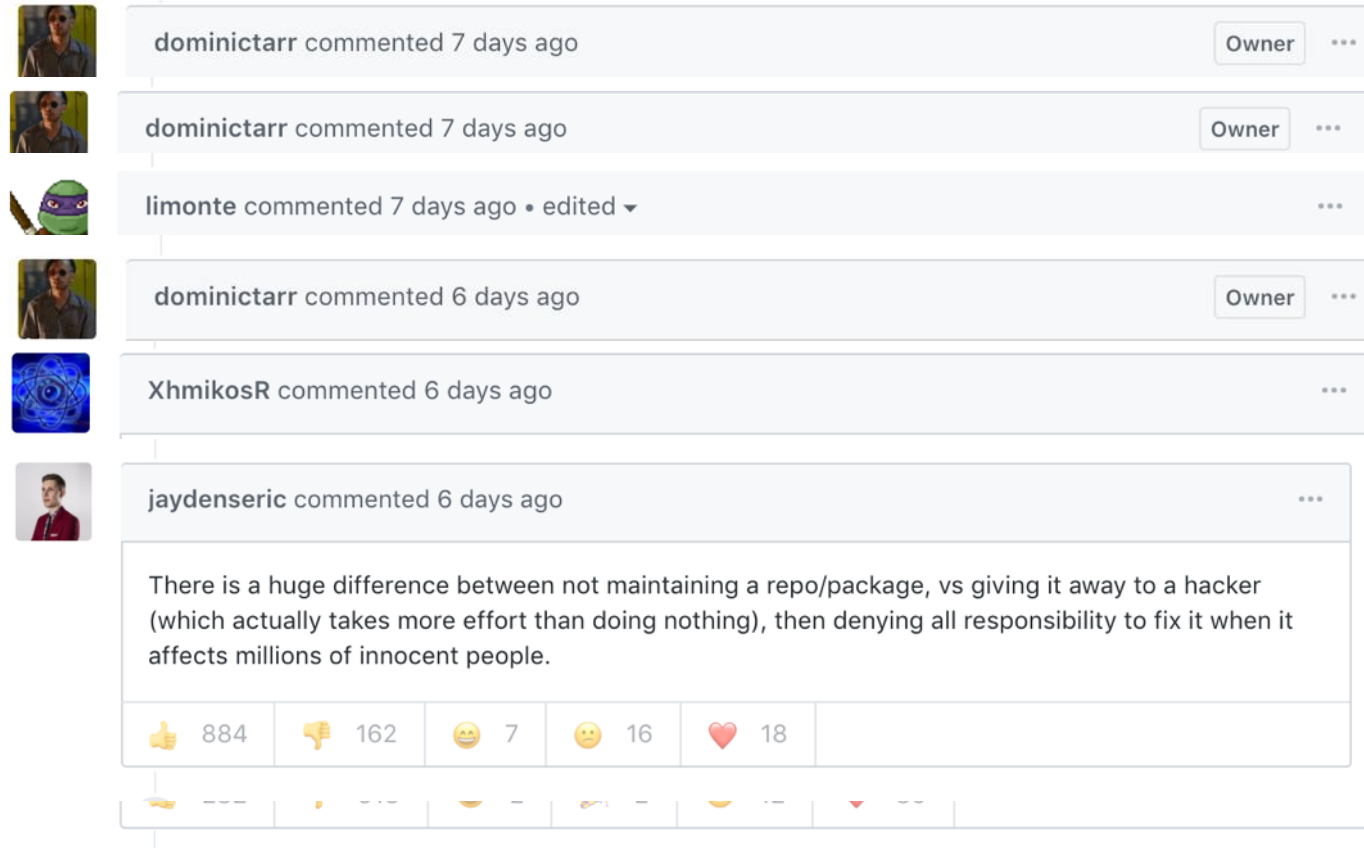
Foresight in AI Chart: Text-to-Image example

Some (of many) example people and use contexts

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

CC BY / Margaret Mitchell / m-mitchell.com

S3D Software and Societal Systems Department

Carnegie Mellon University

Open Source Maintainers

S3D Software and Societal
Systems Department

Carnegie
Mellon
University

Uber self-driving car involved in fatal crash couldn't detect jaywalkers
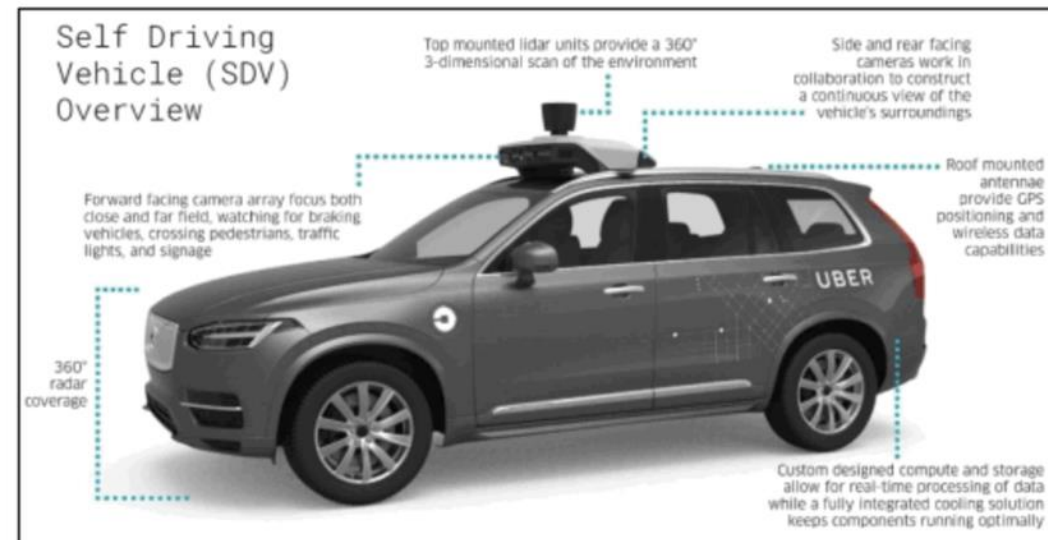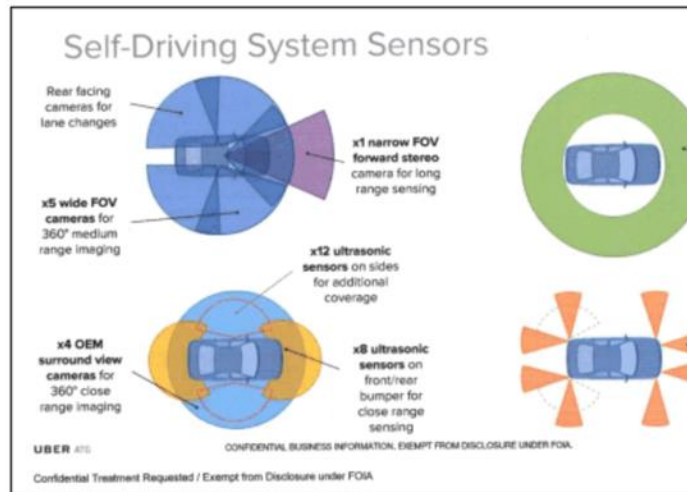
The system had several serious software flaws, the NTSB said.

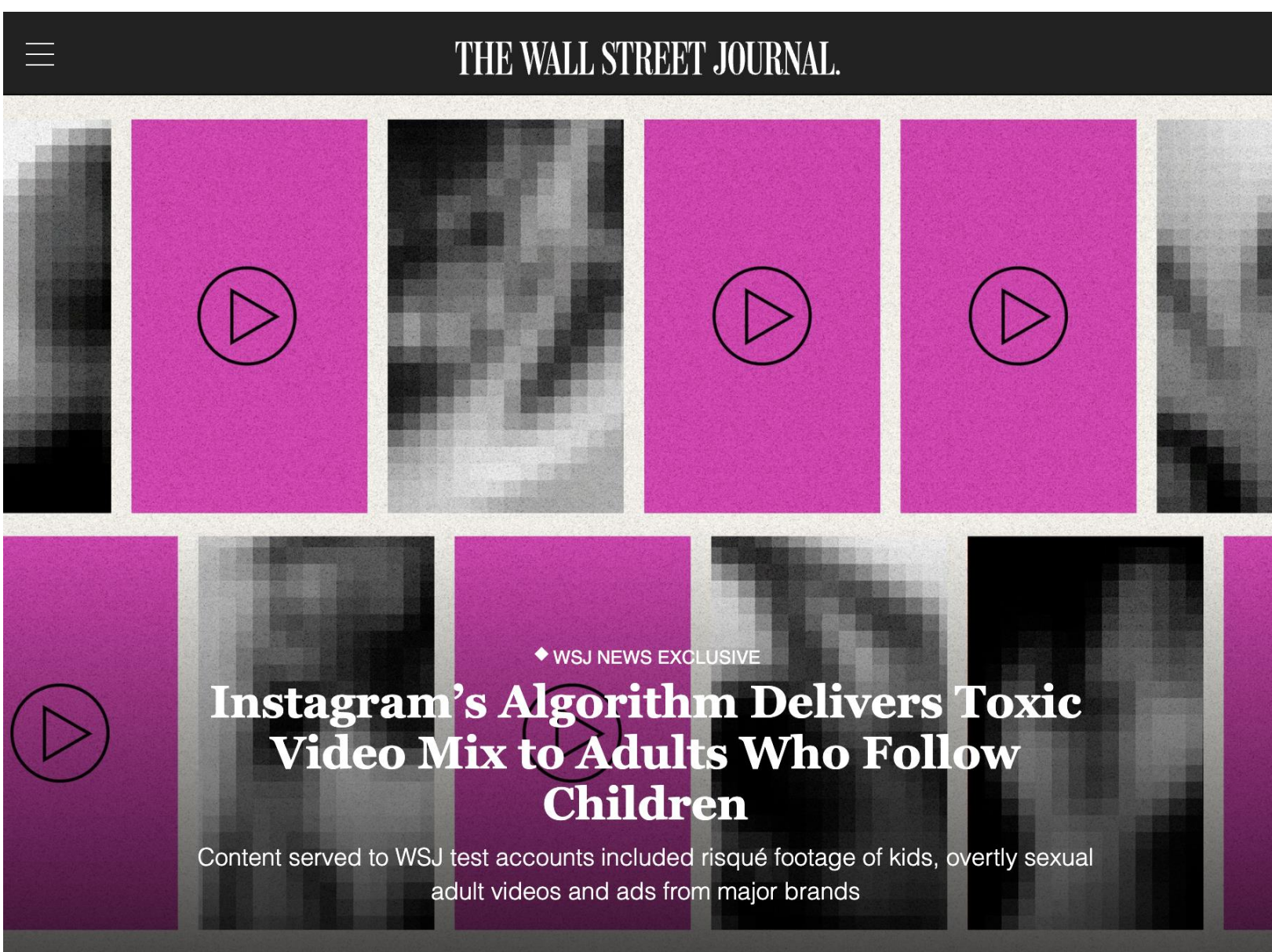Steve Dent, @stevetdent
11.06.19 in Transportation

25
Comments

1131
Shares

THE WALL STREET JOURNAL.

◆WSJ NEWS EXCLUSIVE

Instagram's Algorithm Delivers Toxic Video Mix to Adults Who Follow Children

Content served to WSJ test accounts included risqué footage of kids, overtly sexual adult videos and ads from major brands
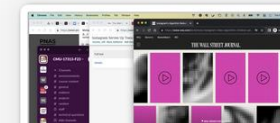
DAISY KORPICS FOR THE WALL STREET JOURNAL

By *Jeff Horwitz* [Follow] *and Katherine Blunt* [Follow]
Nov. 27, 2023 5:30 am ET

Carnegie Mellon University

ZD NET

/ innovation

Home / Innovation / Artificial Intelligence

Microsoft adds new Designer protections following Taylor Swift deepfake debacle

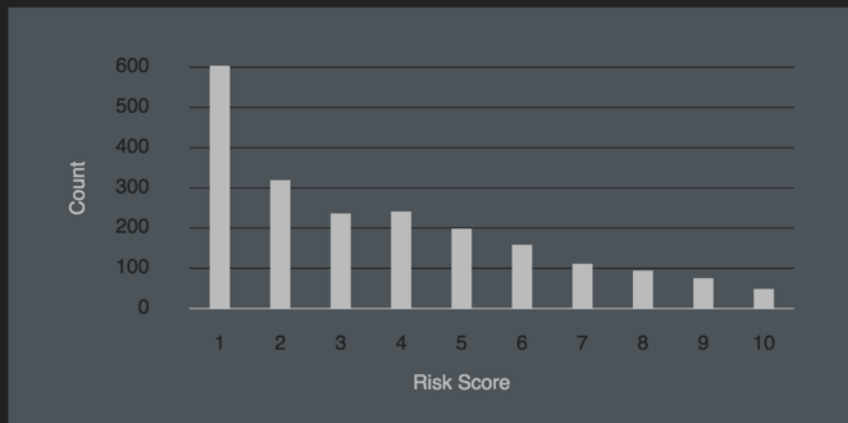The loopholes that allowed the creation of images have been patched.

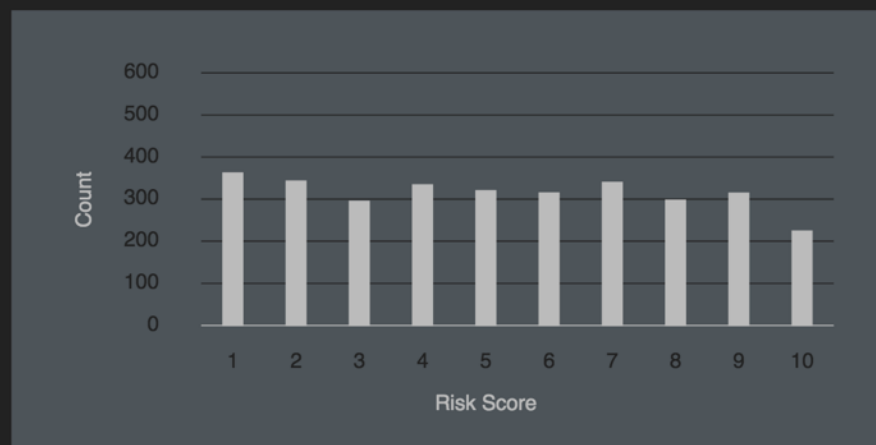Written by **Sabrina Ortiz**, Editor
Jan. 30, 2024 at 11:52 a.m. PT

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

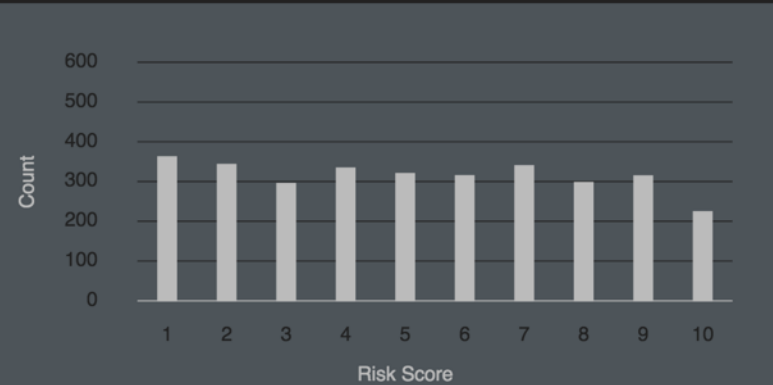Access to health care

Receiving parole
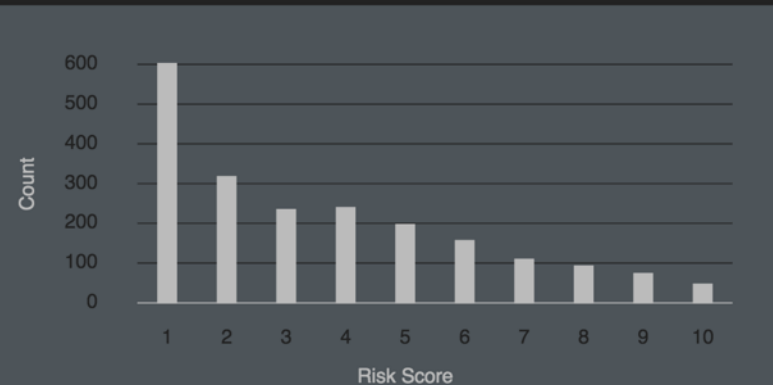
Possibility of Bail

Risk Scores

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.)

S3D Software and Societal Systems Department
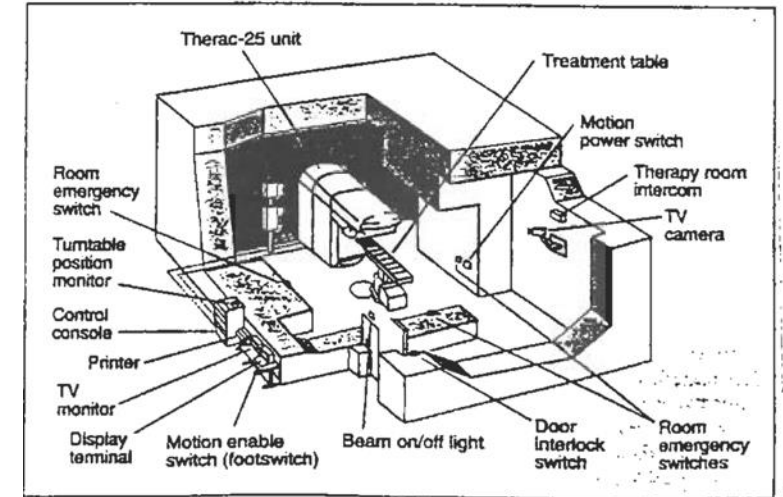
Carnegie Mellon University

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

S3D Software and Societal Systems Department

Carnegie Mellon University

The New York Times

SUBSCRIBE NOW   LOG IN

US edition ˅

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

Make a contribution

Subscribe   Find

News

The

DEV   GUEST

Eng
Opin

Softwa
body

OLGA V. MACK   @

Ho

FE
S

W
et

...t ethics

n
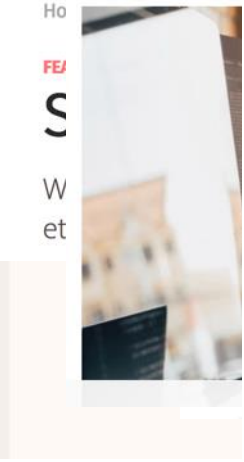
T READ
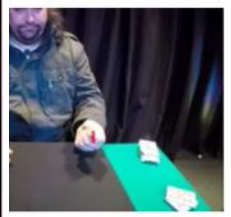
to fool AI with magic

S3D Software and Societal Systems Department

Carnegie Mellon University

Code of Ethics

Association for Computing Machinery

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.

S3D Software and Societal
Systems Department

Carnegie
Mellon
University

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.

S3D Software and Societal Systems Department

Carnegie Mellon University

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.

S3D Software and Societal Systems Department

Carnegie Mellon University

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?

S3D Software and Societal Systems Department

Carnegie Mellon University

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

Carnegie Mellon University

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 Sensitivites are instincts that are often vulnerable to new technologies.

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.	• Artificial scarcity • Urgency signalling • Constant monitoring • Optimizing for screentime	Design engenders calm, balance, safety, pauses and supports circadian rhythms.	High / Low
Attention How and where we focus our attention.	Attention is physiologically drawn, overwhelmed or fragmented.	• Constant context switching • Many undifferentiated choices • Fearful information • No stopping cues (e.g. infinite scroll) • Unnecessary movement	Enabled to bring more focus and mindfulness.	
Sensemaking How we integrate what we sense with what we know.	Information is fear-based, out of context, confusing, or manipulative.	• Facts out of context • Over-personalized filters • Equating virality with credibility • Deceptive authority (ads vs. content)	Enabled to consider, learn, express and feel grounded.	
Decisionmaking How we align our actions with our intentions.	Intentions and agency are not solicited nor supported.	• Avatars to convey authority • Stalking ads and messages • Push content models • Serving preference over intent	Enabled to gain agency, purpose, and mobilization of intent.	
Social Reasoning How we understand and navigate our personal relationships.	Status, relationships or self-image are manipulated.	• Quantified social status • Viral sharing • Implied obligation • Enabling impersonation	Enabled to connect more safely and authentically with others.	
Group Dynamics How we navigate larger groups, status, and shared understanding.	Excluded, divided or mobilized through fear.	• 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.	

[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.

Software and Societal Systems Department

Carnegie Mellon University

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.	• 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

S3D Software and Societal Systems Department

Carnegie Mellon University

GenderMag
https://gendermag.org

Abby Jones[1]
- 28 years old
- Employed as an Accountant
- Lives in Cardiff, Wales

You can edit anything in blue print

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.

[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
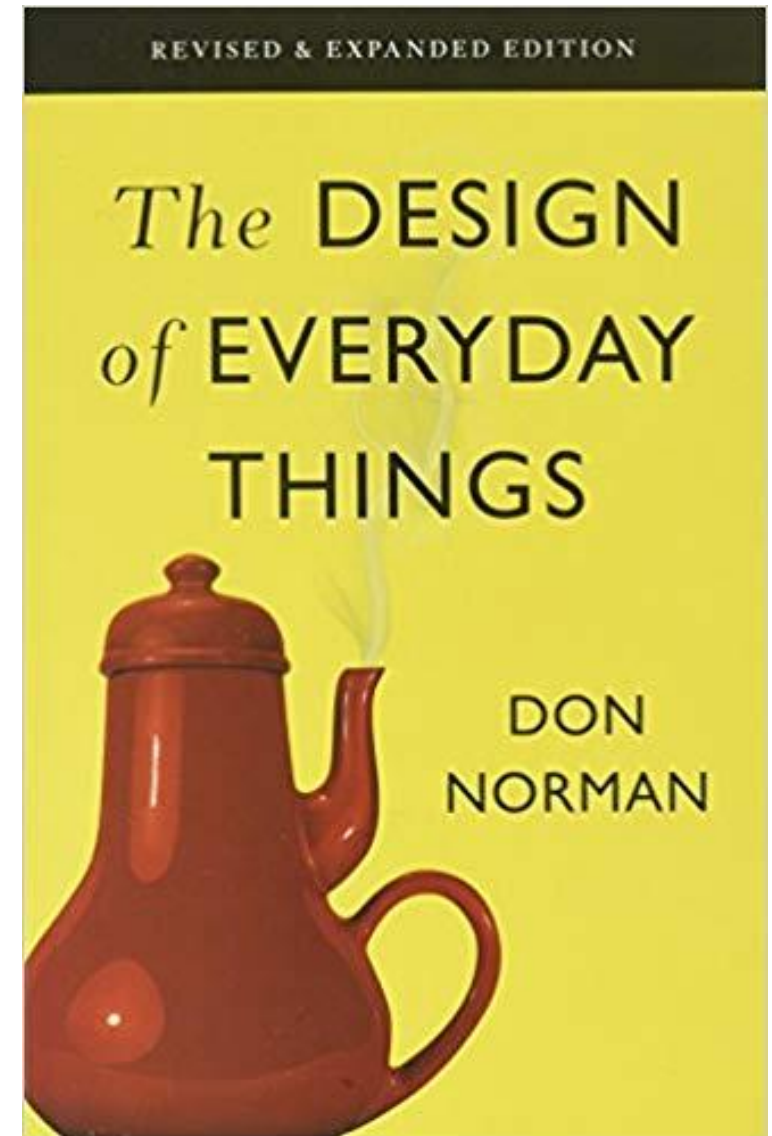
S3D Software and Societal Systems Department

Carnegie Mellon University

GenderMag
https://gendermag.org

S3D Software and Societal Systems Department

Carnegie Mellon University

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
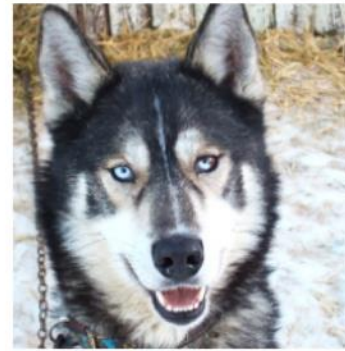
(a) Husky classified as wolf   (b) Explanation
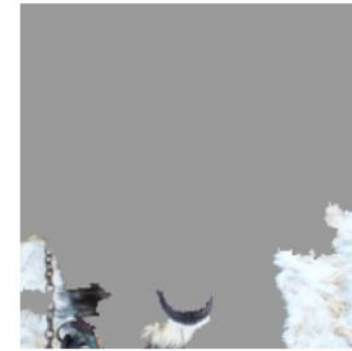
Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

	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

S3D Software and Societal Systems Department

Carnegie Mellon University

Local Interpretable Model-Agnostic Explanations LIME)
https://github.com/marcotcr/lime

Explain "why" to customers
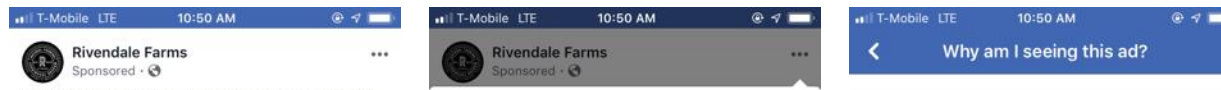
S3D Software and Societal Systems Department

Carnegie Mellon University
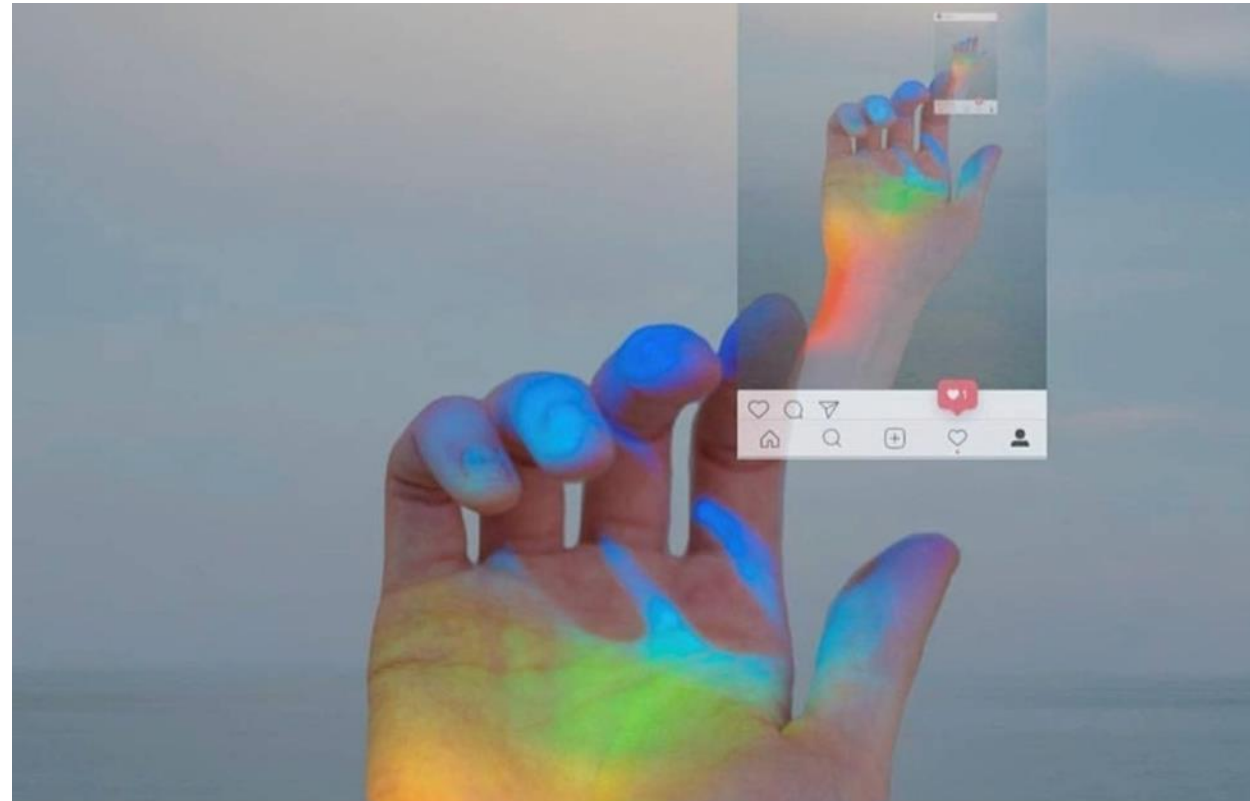
@dovneon

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

Carnegie Mellon University

Anil Dash on how to prevent abuse

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.

S3D Software and Societal Systems Department

Carnegie Mellon University

Deon https://github.com/drivendataorg/deon

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.

S3D Software and Societal Systems Department

Carnegie Mellon University

AI Incident Database

S3D Software and Societal Systems Department

Carnegie Mellon University
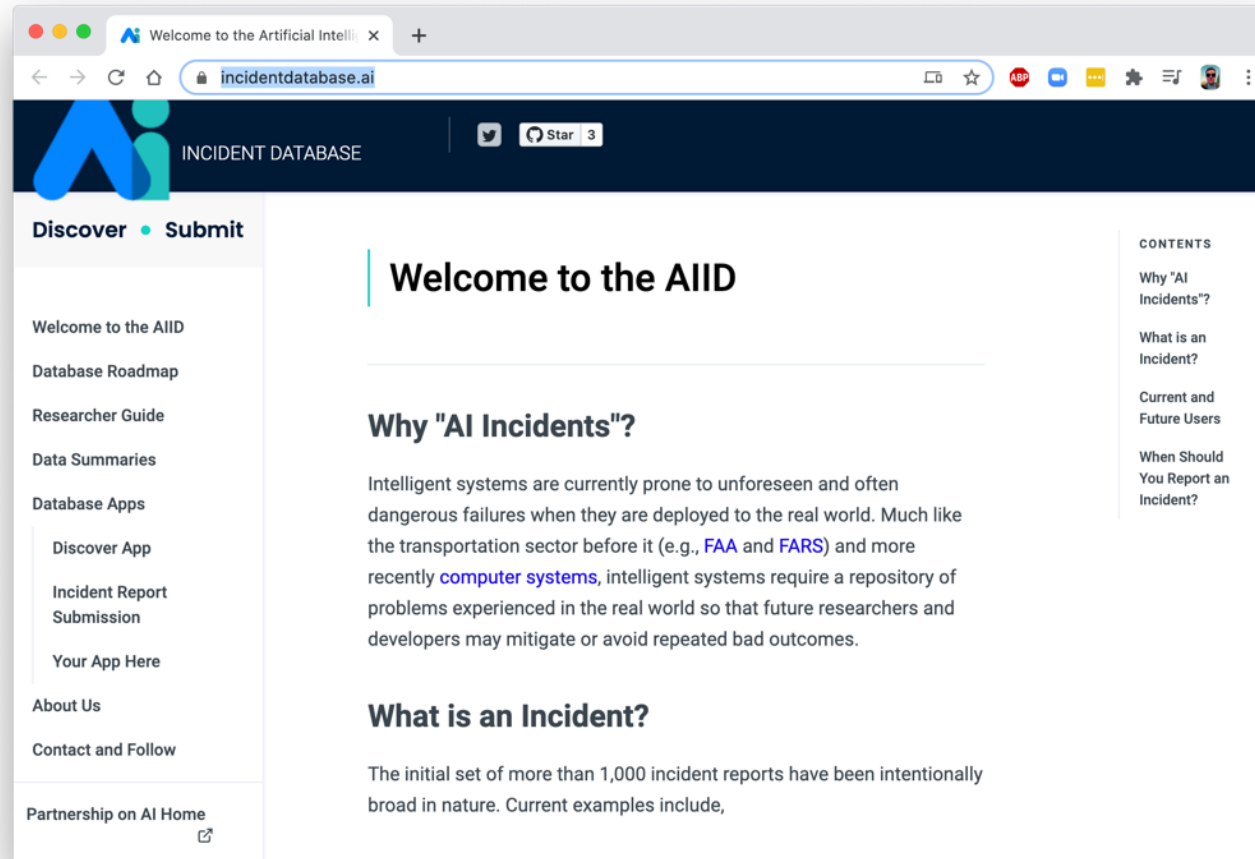
3. Will my software's quality impact the humanity of others?

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

S3D Software and Societal Systems Department

Carnegie Mellon University

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.

S3D Software and Societal Systems Department

Carnegie Mellon University

"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
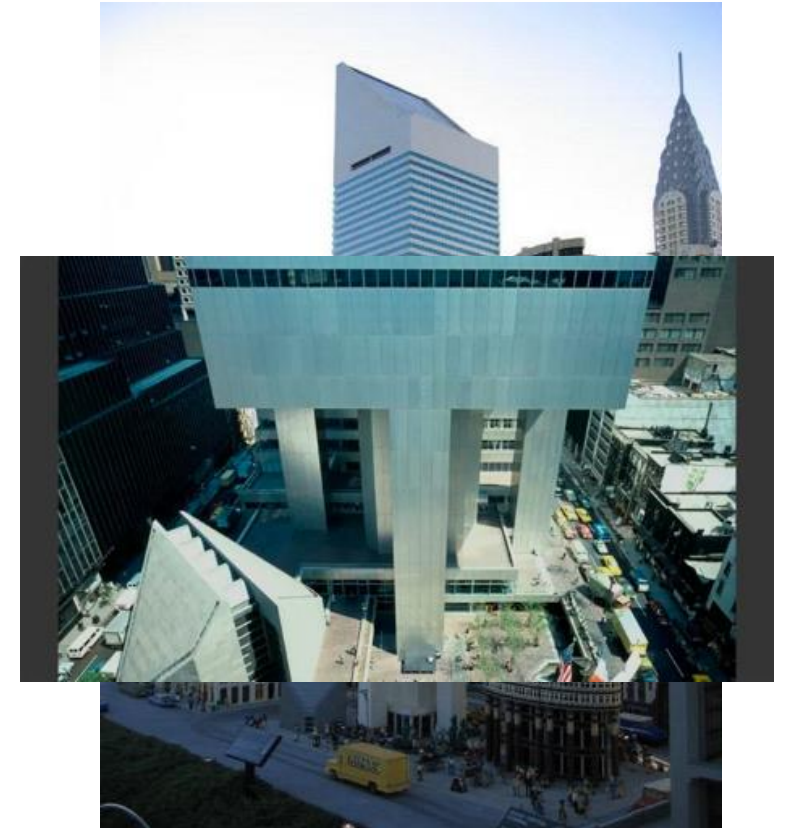
Professional Ethics

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"

S3D Software and Societal Systems Department

Carnegie Mellon University
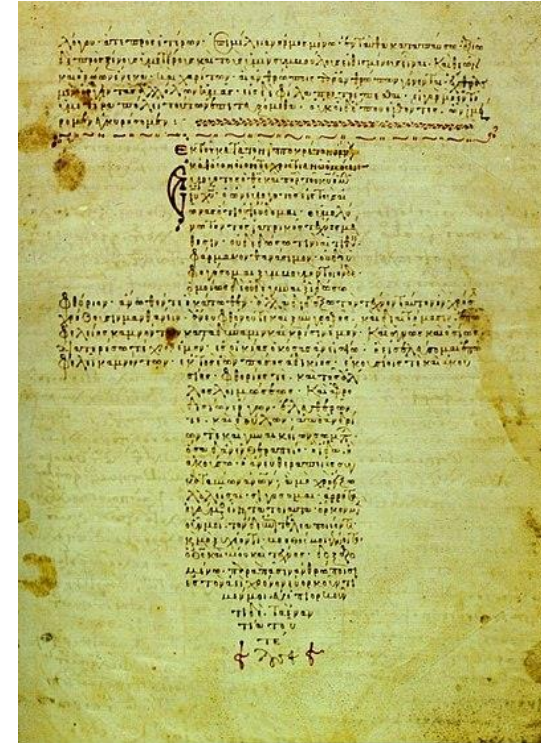
Law -Extrinsic

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

S3D Software and Societal Systems Department

Carnegie Mellon University

Malpractice vs. Negligence

Neglig   **nably**
prude

Malpr
"profe
profes
provid   **ning**
body   the
plainti

YOU CAN'T HAVE MALPRACTICE

IF YOU DON'T HAVE ANY PROFESSIONAL STANDARDS

imgflip.com

S3D Software and Societal Systems Department

Carnegie Mellon University
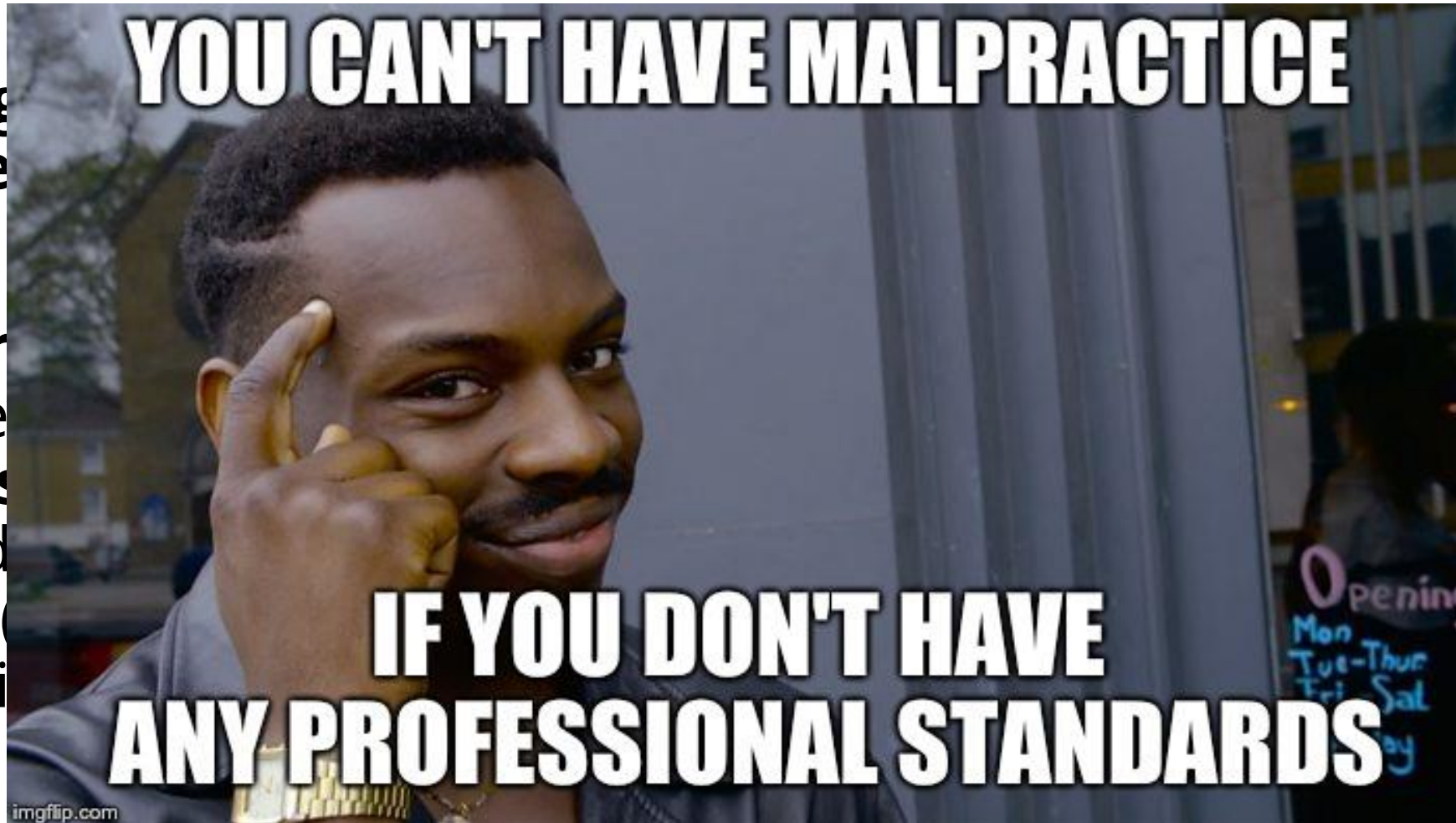
DISCUSSION: What should we do going forward?

Bioengineering Ethics:

- Respect for Autonomy
- Beneficence
- Nonmaleficence
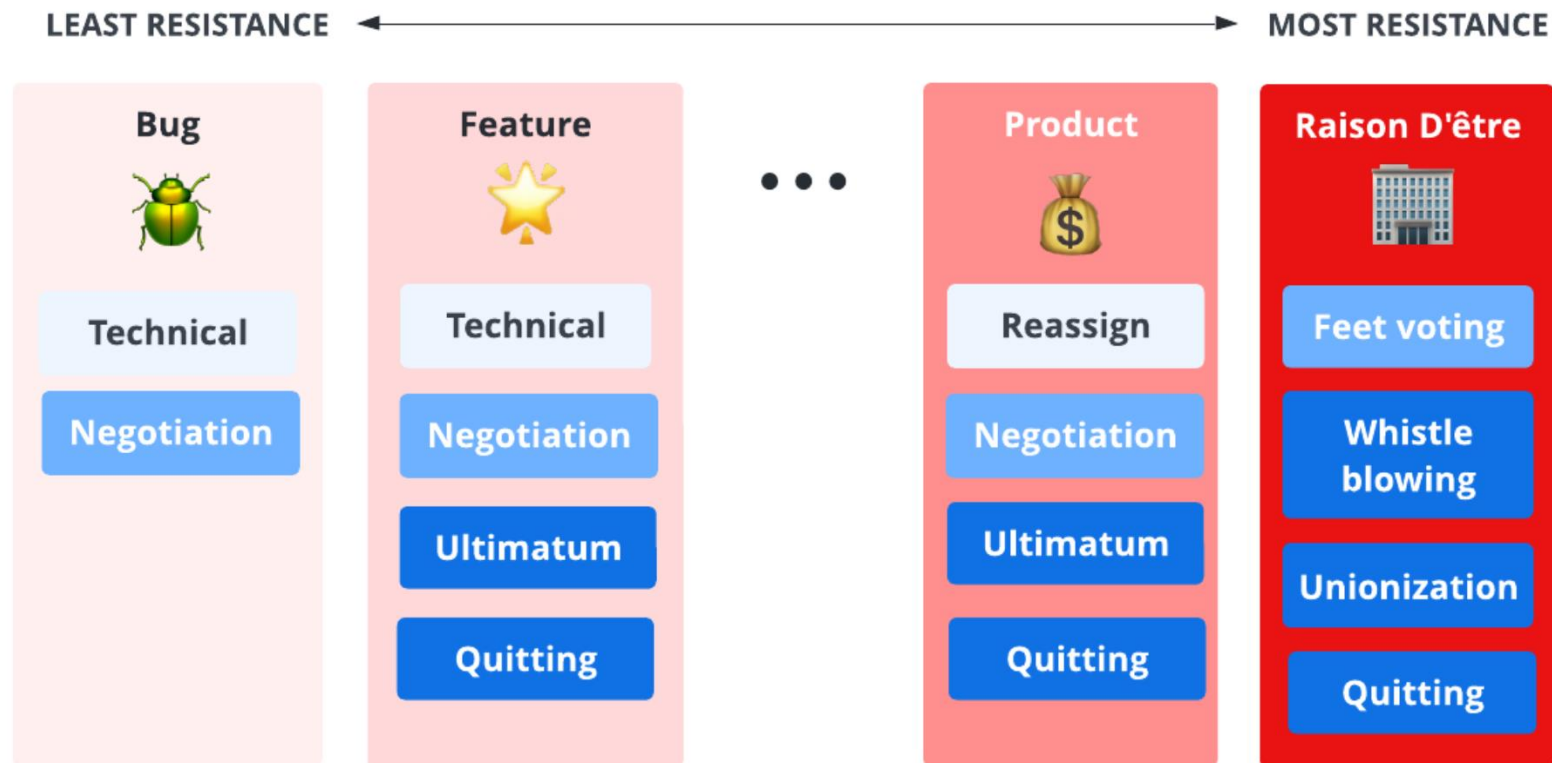- Justice

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

S3D Software and Societal Systems Department

Carnegie Mellon University

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

S3D Software and Societal Systems Department

Carnegie Mellon University

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?

S3D Software and Societal Systems Department

Carnegie Mellon University