

Auditing AI:

The Civil Rights Battle of Our Time

Davi Ottenheimer

whoami (i.e. ISACA since 1997)

- 2005 Campaigned PKI/Token architecture to reduce PCI breaches
- 2006 Patented One-time IoT PIN/secret (“Connected Life” Paranoid)
- 2009 Wrote EKMI → KMIP open standards for key management services
- 2012 Published Securing Virtual Environment (Cloud) Book
- 2013 Started Realities of Securing Big Data (Field-Level Crypto) Book...
- 2017 Started NoSQL Field-Level Crypto DB Product Feature...
- 2018 Created RSAC Humanitarian Service Award (1980s crypto system)
- 2019 Released NoSQL Field-Level Encryption Client-Side (FLECS)

inrupt

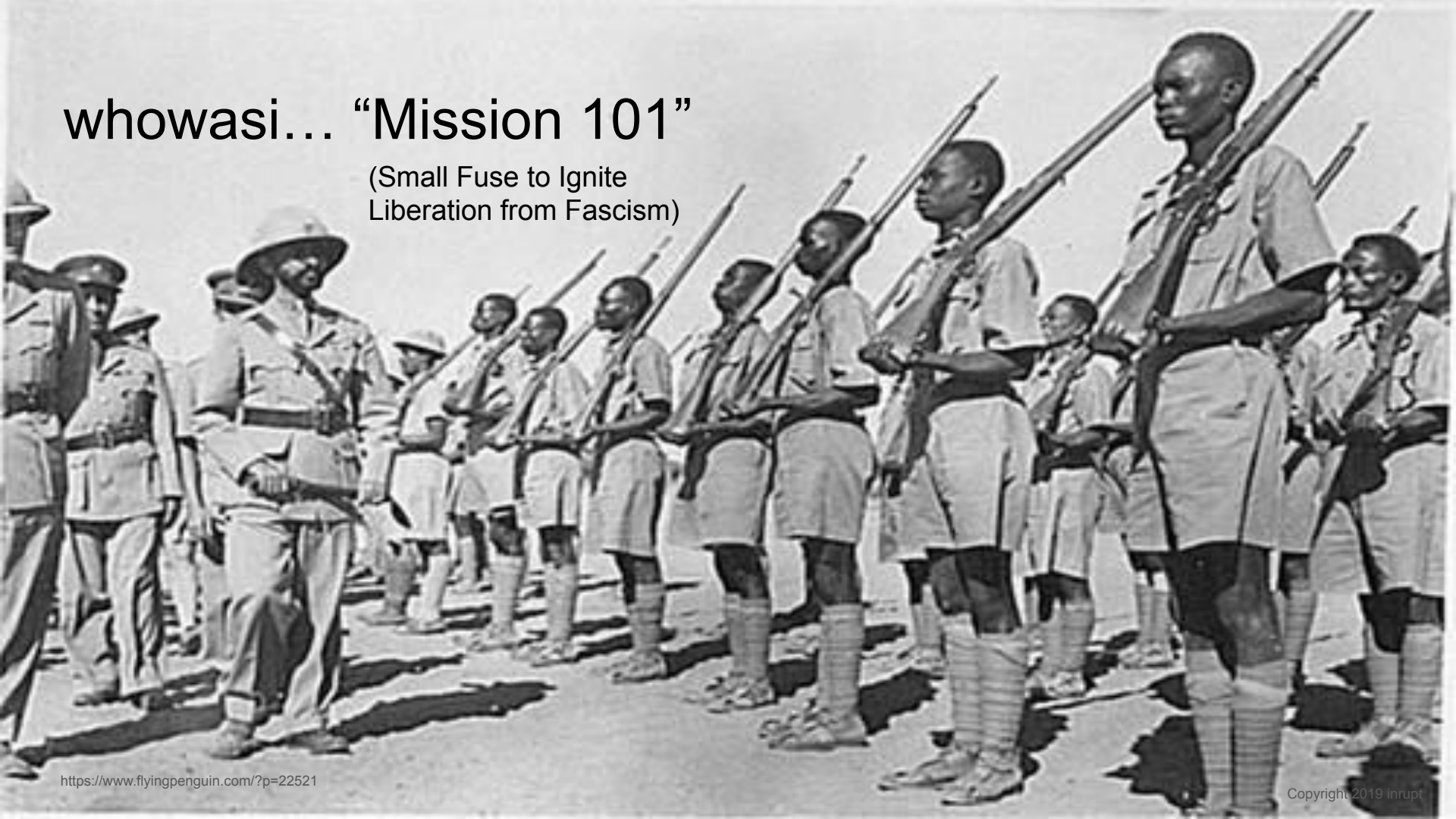


DATA ENTRY
The web is broken, so its founder is taking another stab at it
By Thu-Huong Ho • September 26, 2018



whowasi... “Mission 101”

(Small Fuse to Ignite
Liberation from Fascism)



2012 Presentation: Political Coups via Social Media Mercs

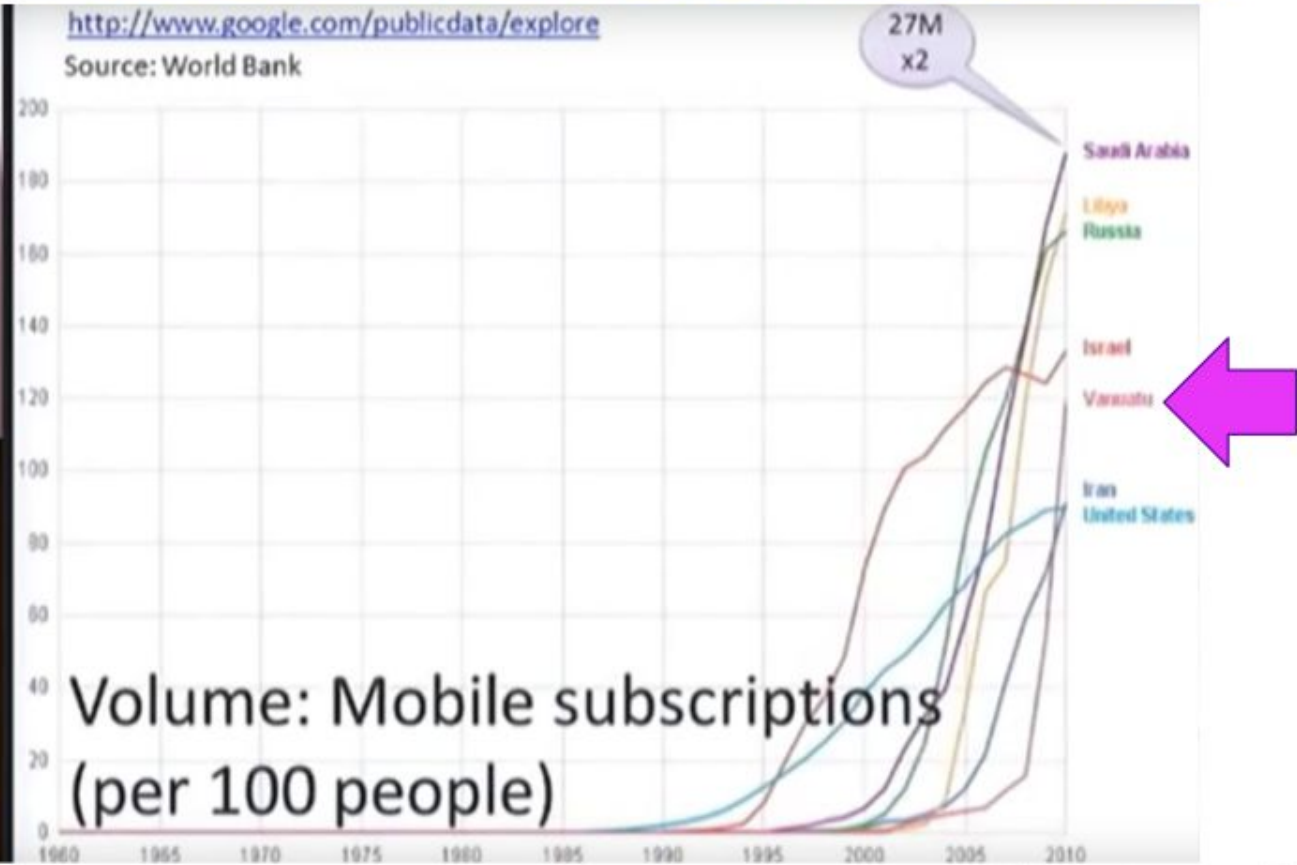


Davi Ottenheimer
Big Data's Fourth V Or
Why We'll Never Find
the Loch Ness Monster

B&SIDES

Las Vegas 2012

MORE VIDEOS



11:39 AM - 31 Jan 2018



<https://twitter.com/daviottenheimer/status/817027735469989888>



Ben Collins ✓

@oneunderscore__

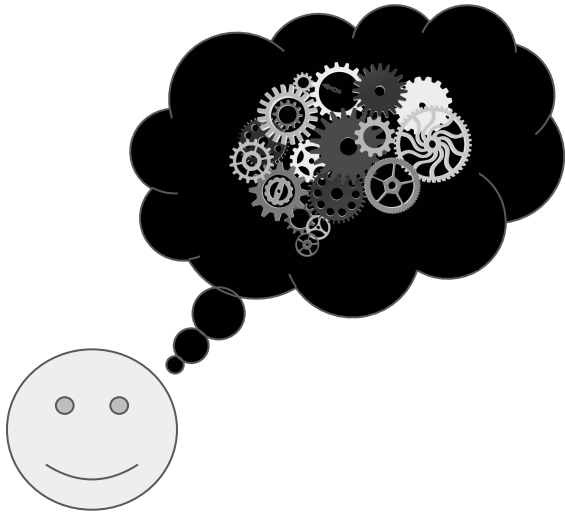
Follow

Top 3 "People Are Saying" posts in Facebook's Trending News section for the Amtrak crash are all absolutely bonkers conspiracy theories.

I follow zero of these people and replicated in Incognito.

It is bananas they have not fixed this problem yet.

The Great Promises of AI



Drive Cars

Translate Language

Pilot Aircraft

Detect Fraud

Detect Malware

Solve Healthcare

And So Much More...!

Rampant racism in decision-making software used by US hospitals

“Hospitals and insurers use the algorithm and others like it to help manage care for about 200 million people in the United States each year...”

Pedestrian deaths increasing year-over-year as cars get smarter

Trusted (security) software
using AI is easily bypassed
by trivial input changes

Some
people very
intentionally
want the
“Fog of War”

Lesson #1: Empathize with your enemy

Lesson #2: Rationality alone will not save us

Lesson #3: There's something beyond one's self

Lesson #4: Maximize efficiency

Lesson #5: Proportionality should be a guideline in war

Lesson #6: Get the data

Lesson #7: Belief and seeing are both often wrong

Lesson #8: Be prepared to reexamine your reasoning

Lesson #9: In order to do good, you may have to engage in evil

Lesson #10: Never say never

Lesson #11: You can't change human nature

<https://archive.org/details/TheFogOfWarElevenLessonsFromTheLifeOfRobertS.Mcnamara>

HOW

R

WE

SUPPOSED TO AUDIT?

Like Asking Can I Audit Your Baby's Intelligence?



Like Asking Can I Audit Your Baby's Likelihood of Being Racist?



Uncomfortable *Truth* About Learning...All Around Us



Jefferson
President of the
Confederate States

...

Beauregard
First General of
Confederate States,
"South's Paladin"

...

"I thought the KKK was ok until I learned they smoked pot"

Judgeship nominee denies he is racist

By John McCaslin
THE WASHINGTON TIMES

President Reagan's nominee to a federal judgeship in Alabama told Senate Judiciary Committee members yesterday he is "not the Jeff Sessions my detractors have tried to create."

"I am not a racist. I am not insensitive to blacks," Jefferson B. Sessions III, the 38-year-old U.S. attorney for south Alabama, said at an unprecedented fourth confirmation hearing. "I have done my job with integrity, equality and fairness for all."

The committee earlier postponed a vote on the nomination to give Mr.

said Mr. Kennedy. "I don't intend to repeat all of the racial remarks which Mr. Sessions is alleged to have made, many of which he has acknowledged in his own testimony."

Mr. Sessions, however, took issue with that statement. "I did not admit to any insensitive statements then, and I do not now," he said.

Testifying under oath, Mr. Sessions denied having ever described, as one witness claimed, the National Association for the Advancement of Colored People and the National Council of Churches as "un-American."

He denied agreeing with another person's statement that a white civil rights lawyer in Mobile, Ala., was a



Jefferson B. Sessions III

convinced that, after a fair and full

Sessions the Third. Born 1946
Father born 1916?
(America First = Rise of KKK)

Uncomfortable *Truth* About Learning...All Around Us

1740 South Carolina Ban on Teaching Slaves to Write

1758 Georgia Ban on Teaching Slaves to Write

1833 Alabama Set Fines for Educating Slaves

1836 North Carolina Ban on Education of Blacks

1841 Mississippi Required Educated Black Freemen to Leave State

[...]

1959 Virginia Shuts Public Schools and Gives *Vouchers for Whites Only*

Uncomfortable *Truth* About Learning...All Around Us

Voucher Schools Championed By Betsy DeVos Can Teach Whatever They Want. Turns Out They Teach Lies.

NELSON MANDELA WAS MARXIST

SATAN CREATED PSYCHOLOGY

GOD WANTS ONLY MODEST WOMEN

THE "WAR BETWEEN THE STATES"

BLACK SUPREMACIST ACTIVISTS

2018 Microsoft 10-Q: Learning Systems May Harm

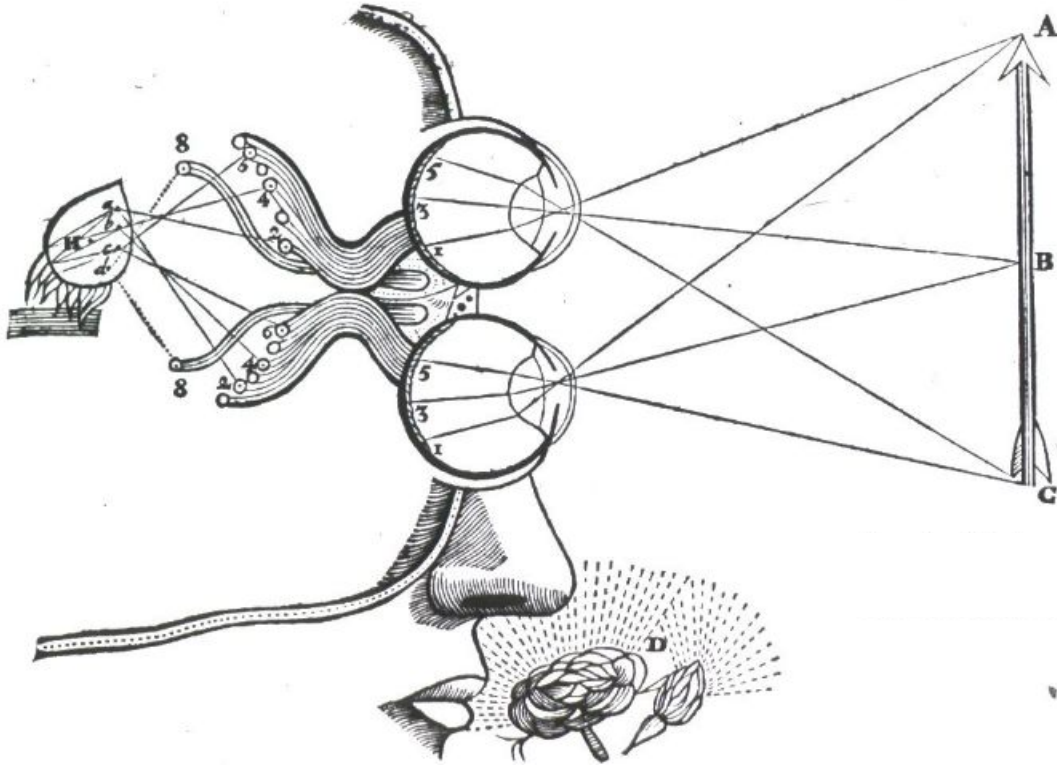
Issues in the **use of AI in our offerings may result in reputational harm or liability**. We are building AI into many of our offerings and we expect this element of our business to grow. We envision a future in which AI operating in our devices, applications, and the cloud helps our customers.... As with many disruptive innovations, AI presents risks and challenges that could affect its adoption, and therefore our business. **AI algorithms may be flawed. Datasets may be insufficient or contain biased information. Inappropriate or controversial data practices** by Microsoft or others could impair the acceptance of AI solutions. These deficiencies could undermine the decisions, predictions, or analysis AI applications produce, subjecting us to competitive harm, legal liability, and brand or reputational harm. Some **AI scenarios present ethical issues**. If we enable or offer AI solutions that are controversial because of their **impact on human rights, privacy, employment, or other social issues**, we may experience brand or reputational harm.

Could We Audit for a “Criminal Brain” in Machines?

Stop the Obvious Monster

A SEC Compliance Examiner breached a US government investigations database to “dangle” stolen intelligence in front of an investigation target as their incentive to hire him as their corporate compliance officer.

Mary Shelley's 1816 "Frankenstein" Warned Us



HENRY: It's a perfectly good brain, doctor. Well, you ought to know. It came from your own laboratory.

WALDMAN: The brain that was stolen from my laboratory was a

criminal brain.

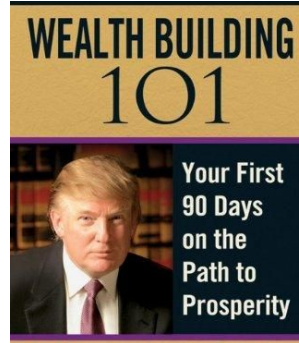
Mary Wollstonecraft (Mary Shelley's Mother) Too

- A Vindication of the Rights of Man (1790)
 - Unequal society founded on passivity of women
 - Rationality, unlike ancestral traditions or dogma, abolishes slavery
- A Vindication of the Rights of Woman (1792)
 - Human limitations are a result of deficient education
 - Middle-class “most natural state”
 - Equality of sexes

‘I attribute [these problems] to a **false system of education**, gathered from the books written on this subject by men, who, **considering females rather as women than human creatures**, have been more anxious to make them alluring mistresses than affectionate wives and rational mothers...’

Machines Accelerate False Systems of Education

- ***Pancake Robot*** learns to throw as high as possible to “avoid” ground
- ***Driving Robot*** goes in reverse and impacts body to avoid touching bumpers
- ***Tic-tac-toe Robot*** makes distant moves to **cause opponent memory exhaustion and forfeit**



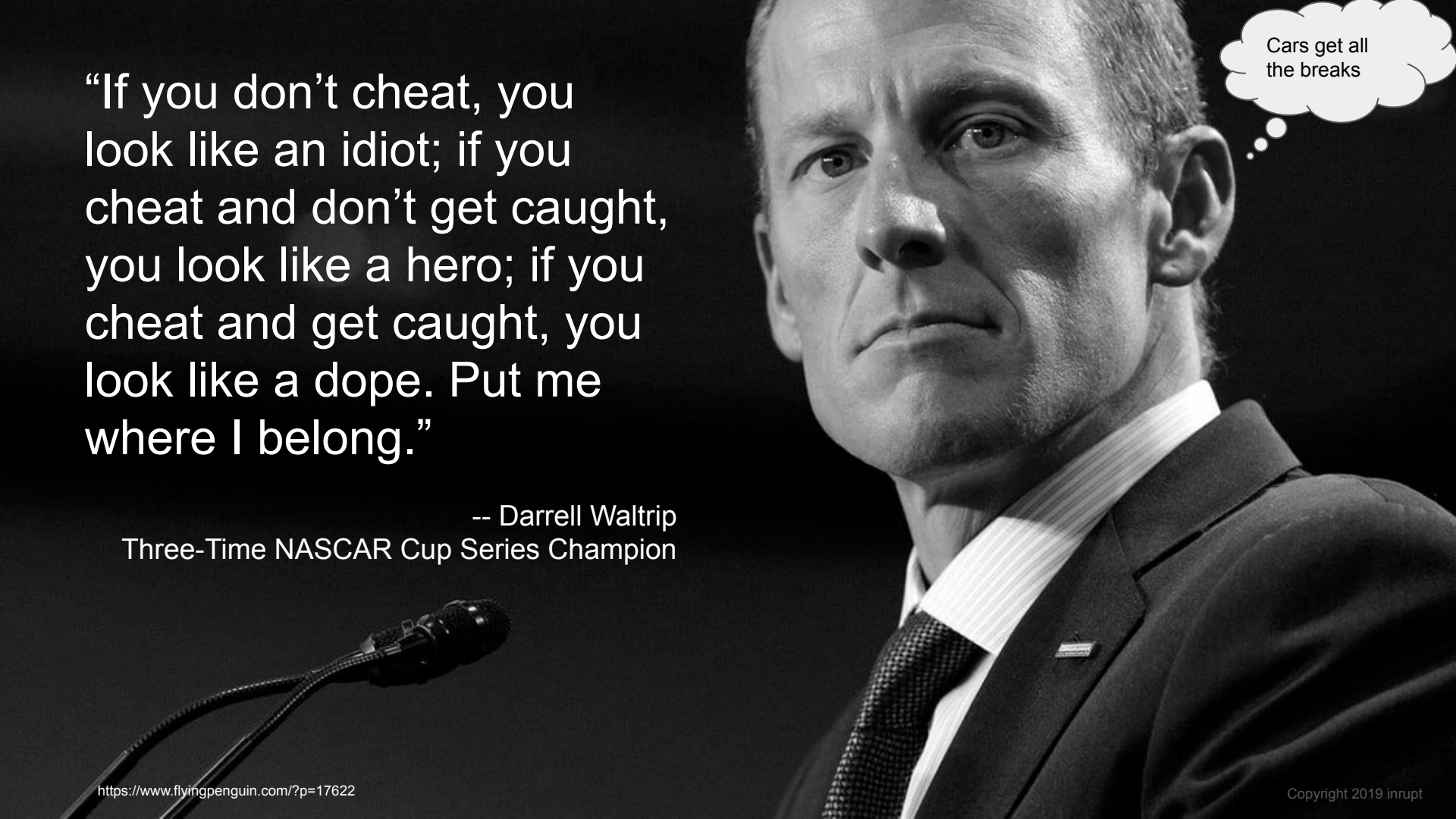
6,000 Victims of 101 Scam

Lawsuits charge students were told to lie about their income to persuade credit card companies to increase credit limits for payment to Trump U.

Internal Trump U. document titled “**surefire script to more purchasing power**” told students to deliberately lie about actual income on credit applications, use “projected income” from imaginary future business ideas.

“If you don’t cheat, you look like an idiot; if you cheat and don’t get caught, you look like a hero; if you cheat and get caught, you look like a dope. Put me where I belong.”

-- Darrell Waltrip
Three-Time NASCAR Cup Series Champion



Cars get all the breaks

Who Believed Armstrong Was Really That Good?

What if Intelligence Reveals What We Can't Believe?

Class-based (Racist) Healthcare System

“...because the algorithm assigned people to high-risk categories on the basis of costs, those biases were passed on in its results: ***black people had to be sicker than white people before being referred for additional help***. Only 17.7% of patients that the algorithm assigned to receive extra care were black. The researchers calculate that the proportion would be 46.5% if the algorithm were unbiased.”

What if Intelligence Reveals What We Can't Believe?

Class-based (Racist) Traffic Management

“Auto campaigners lobbied police to publicly shame transgressors by whistling or shouting at them — and even carrying women back to the sidewalk — instead of quietly reprimanding or fining them. They staged safety campaigns in which actors dressed in 19th-century garb, or as clowns, were hired to cross the street illegally, signifying that the practice was outdated and foolish. In a 1924 New York safety campaign, ***a clown was marched in front of a slow-moving Model T and rammed repeatedly.***”

What if Intelligence Reveals What We Can't Believe?

How would you audit
for driverless cars
***ACCELERATING a
history of violence
against pedestrians
(non-whites)?***

Jacksonville's "Jaywalking" Enforcement Is Very, Very Racist

By Angie Schmitt | Nov 16, 2017 | 56



Jacksonville, Florida has some of the nation's most dangerous roads for pedestrians. The city's police have cynically exploited a genuine public safety threat to use "jaywalking" as a pretext to stop and search black residents. Image: Florida Times-Union

“Crash-Avoiding Tech”

=

Higher Levels of Death?

Traffic Deaths Last Year Reached 28 Year Highs

2019 News Headline:

U.S. Traffic Deaths Drop as
Agency Points to
Crash-Avoiding Tech

Article Text:

“...deaths of pedestrians and bicyclists each rose to ***28-year highs last year***. The number of people killed in large trucks also rose 0.8% to 885, ***the most since 1988***”

A Predictable Result of Crash-Avoiding Tech ?

An estimated 5,997 pedestrian fatalities occurred during 2016, compared with 5,376 in 2015 and 4,910 in 2014.

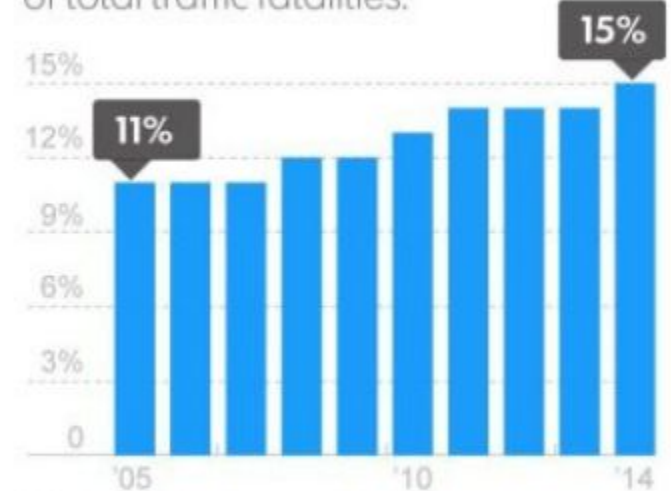
*2016 estimate based on preliminary data



Source: GHSA

PEDESTRIAN DEATHS

Pedestrian deaths as a percentage of total traffic fatalities:



SOURCE: National Highway Traffic Safety Administration
Jim Sergent, USA TODAY

USA TODAY

Is There a “Crash-Avoiding Tech” Engineer Code of Ethics?

SF Day-Time: Uber Driverless Car Trials Caught Running Red-Lights and Cross-Walks

Remember: "...marched in front of a slow-moving Model T and rammed repeatedly"?



davi ((())) 德海 @daviottenheimer · 14 Dec 2016

any comment @sfpd @sfmta @walksf on driverless [redacted] running red lights, ignoring pedestrians in crosswalk? [twitter.com/JoeBeOne/statu...](https://twitter.com/JoeBeOne/status/801111111111111111)



1 2 6

SF Night-Time: Uber Kills Pedestrians

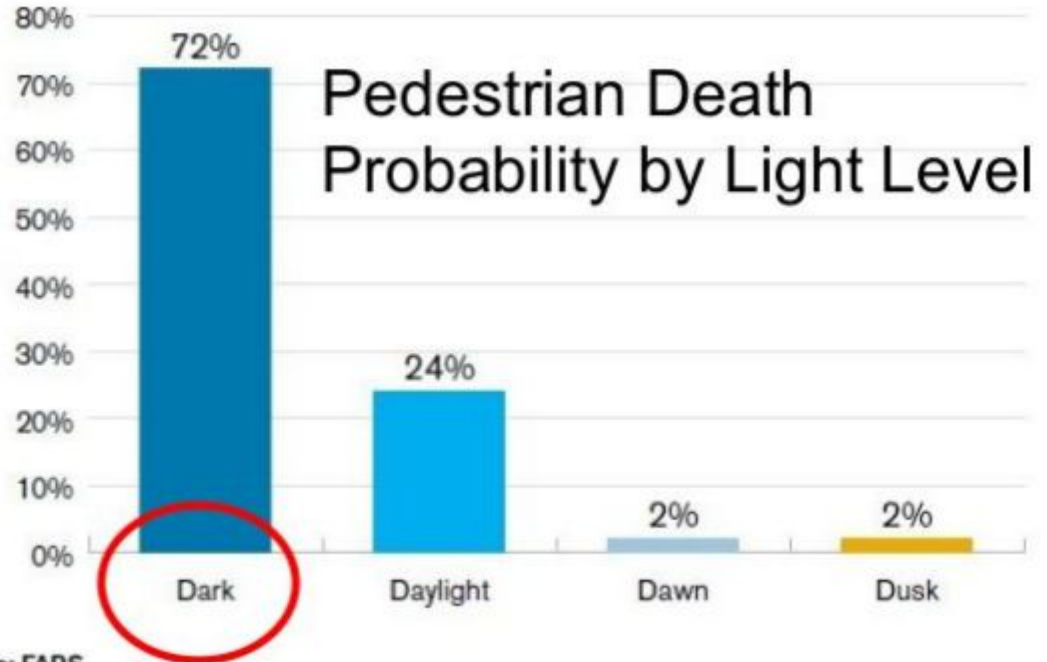


“41% increase in hit and runs in [London], says City Hall”

Commercial “Reasoning” of Driverless Cars Could Diminish Likelihood of Humanity:

- Maximize Rides/Hour
- Maintain Availability
- Avoid Dwell Time
- Cost of Business if Caught

Arizona in the Dark: Driverless Uber Kills Pedestrian



Source: FARS

What Do You See?

2021-10-25 21:55:47
N50.425041 E30.420975
55K m/H



Is There a “Crash-Avoiding Tech” Engineer Code of Ethics?

Tesla 2016 “Autopilot”



Chose to kill human
because 'overhead sign'
(more likely a moving bridge)

NTSB Report: Manufacturers Responsible

Tesla 2017 “Autopilot”



Venkat Viswanathan

@venkvis

Follow

██████████ autopilot camera misreads 101 sign as 105 speed limit at 87/101 junction San Jose. Reproduced every day this week.

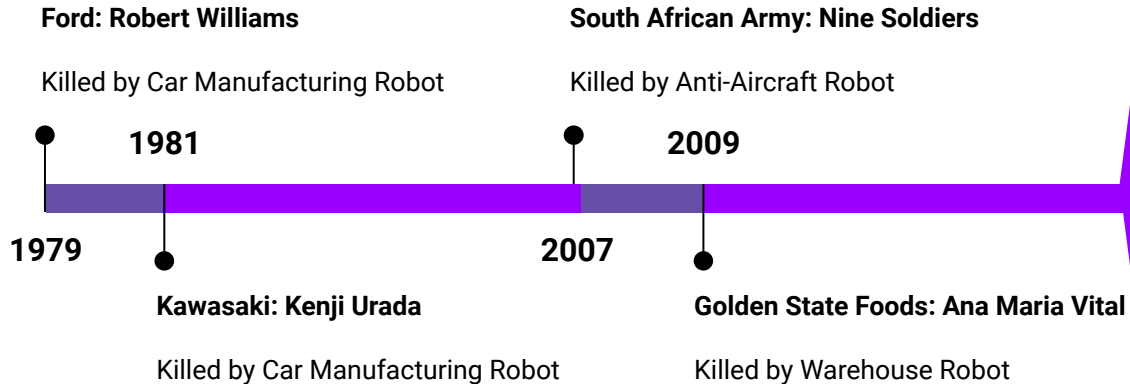


Tesla 2018 “Autopilot”

“Autopilot function was engaged for the last 18 minutes and 55 seconds of Huang’s drive that Friday morning...following a car until seconds before the crash. But the car either changed lanes or exited and once there were no vehicles in front of the Tesla, it began to accelerate. ‘At 3 seconds prior to the crash and up to the time of impact with the crash attenuator, the Tesla’s speed increased from 62 to 70.8 mph, with **no pre-crash braking or evasive steering movement detected,**’ the report stated.”



Accelerating Deaths from Robots



2015-2019

- **2015 VW: Anonymous**
Killed by Car Manufacturing Robot
- **2015 SKH Metals: Ramji Lal**
Killed by Welding Robot
- **2016 Dallas: Micah Johnson**
Killed by Bomb Defuser Robot
- **2016 Ajin USA: Regina Elsea**
Killed by Car Manufacturing Robot
- **2016 Tesla: Joshua Brown**
Killed by Driverless Car
- **2017 VIM: Wanda Holbrook**
Killed by Car Manufacturing Robot
- **2018 Uber: Elaine Herzberg**
Killed by Driverless Car
- **2019 Boeing: 346 Passengers**
Killed by Pilotless Planes

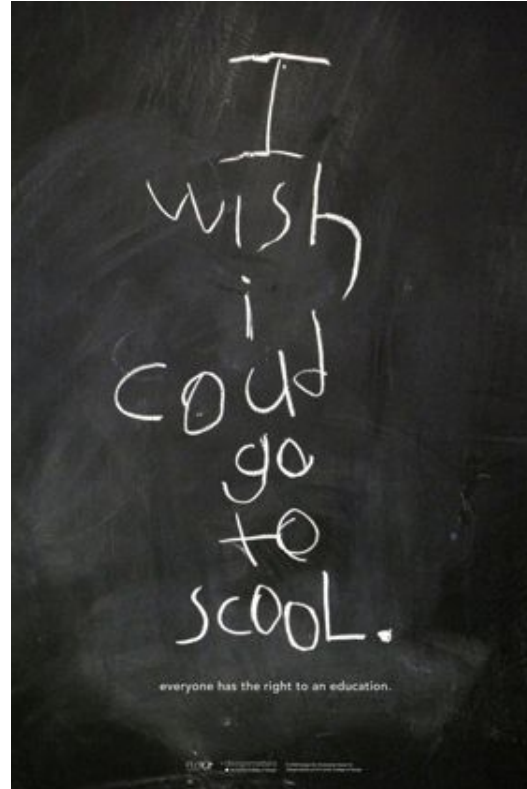
Factors of Auditability (Machine Education Quality)

Auditors

- Independence
- Value Systems
- Actionable Results

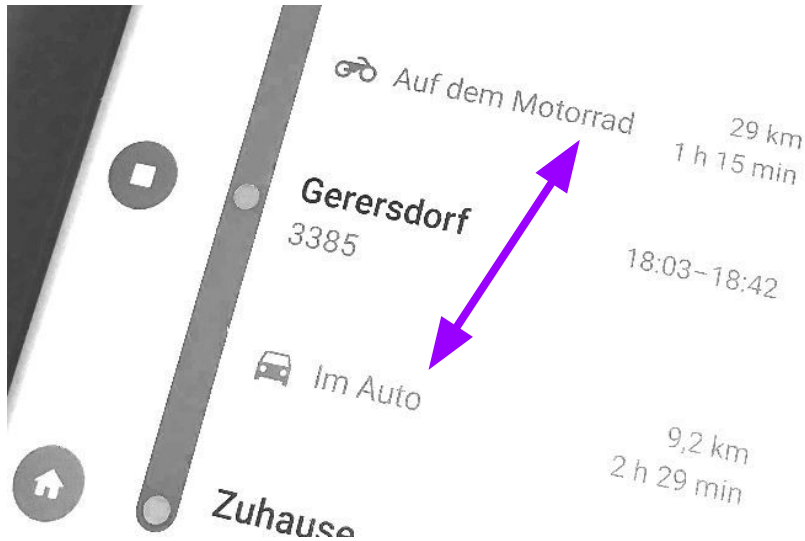
Learning Systems

- Lack of Transparency
- Lack of Standards
- Overconfidence



Lack of Transparency (Consent)

Bias During Design



Researchers: "Classified as a rifle from every angle!"
Reality: It's still classified as a turtle. Rifle not listed



Lack of Transparency (Consent)

Bias During Design

“...intruders can sneak around undetected by holding a small cardboard plate in front of their body aimed towards the surveillance camera.”

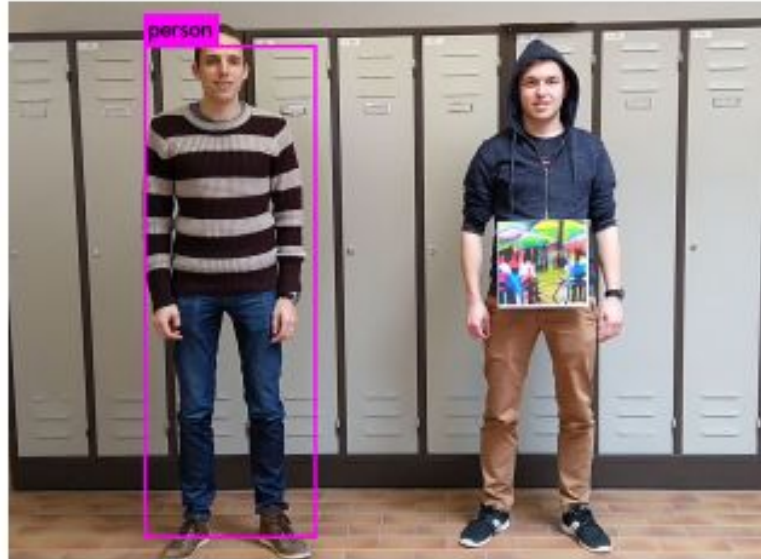
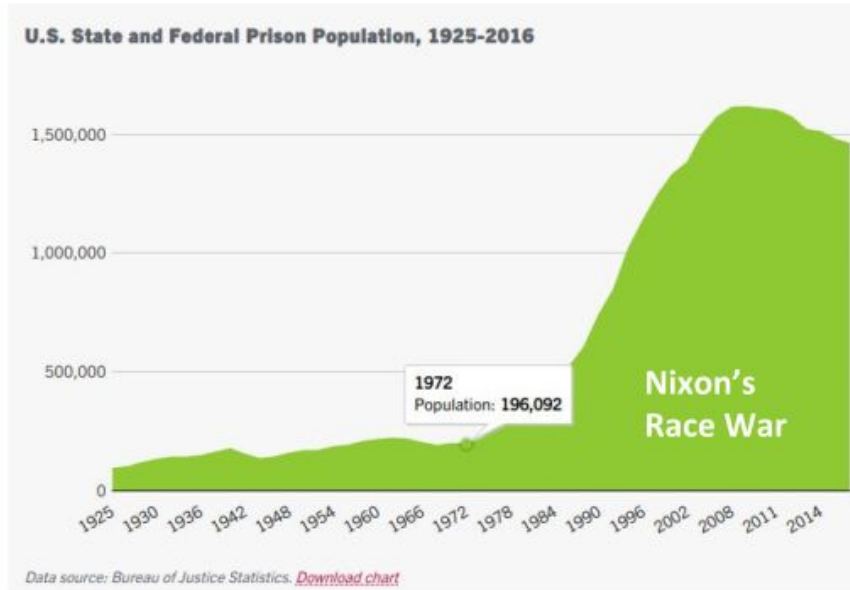


Figure 1: We create an adversarial patch that is successfully able to hide persons from a person detector. Left: The person without a patch is successfully detected. Right: The person holding the patch is ignored.

Lack of Transparency (Consent)

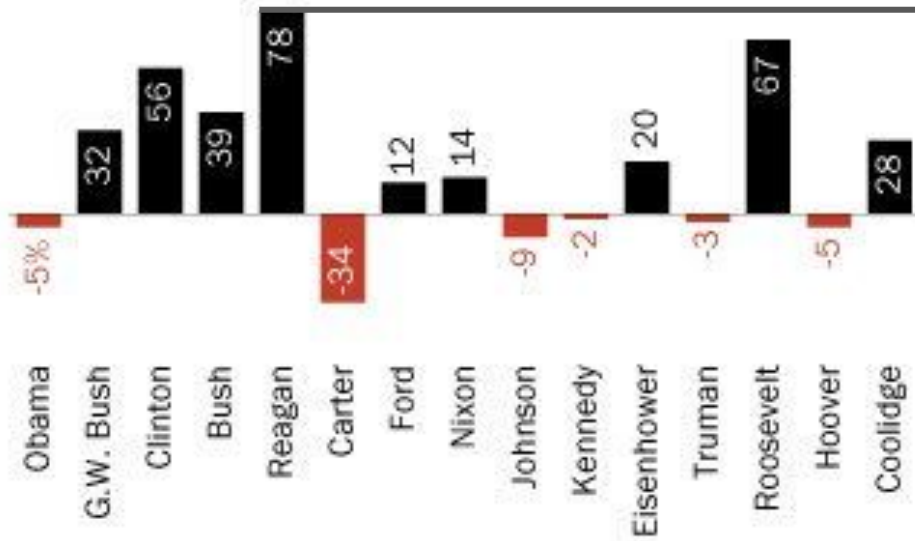
Model Training With Biased Data: Recidivism



“You understand what I'm saying? We knew we couldn't make it illegal to be either against the war or black, but by getting the public to associate the hippies with marijuana and blacks with heroin. And then criminalizing both heavily, we could disrupt those communities. We could **arrest their leaders, raid their homes, break up their meetings, and vilify them night after night** on the evening news. Did we know ***we were lying*** about the drugs? Of course we did.”

Lack of Transparency (Consent)

Model Training With Biased Data: Recidivism



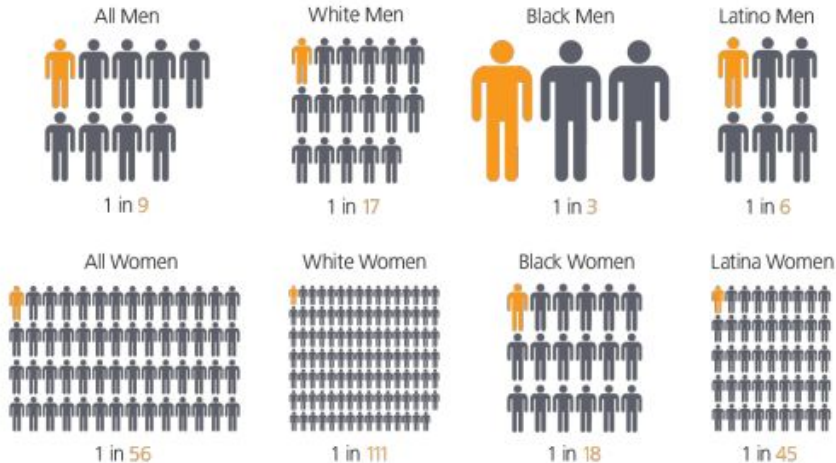
Reagan's 1982 enactment of Nixon's "War" on Drugs (targeting blacks) ran far higher incarceration rates than even *Roosevelt's internment camps during WWII*

Does AI make it worse?

Lack of Transparency (Consent)

Model Training With Biased Data: Recidivism

Lifetime Likelihood of Imprisonment of U.S. Residents Born in 2001



“Blacks **falsely labeled** [by ML as] future criminals at almost twice the rate of white defendants”

“...compared predicted to actual recidivism: scores **wrong 40%** of the time and **biased against black** defendants.”

Source: Bonczar, T. (2003). *Prevalence of Imprisonment in the U.S. Population, 1974-2001*. Washington, DC: Bureau of Justice Statistics.



Lack of Transparency (Consent)

Model Training With Biased Data: Deliveries

“They lost the social connections, the cultural connections, and the communities that were forever torn apart. They bore the consequences. That’s why ‘urban renewal’ gets called **‘Negro removal’; about two-thirds of the people who were displaced were minorities.** [...] By 1965, nearly 800 cities around the country—located across almost every state—were participating in urban renewal.”

Lack of Transparency (Consent)

Model Training With Biased Data: Deliveries

'Berman says ethnic composition of neighborhoods isn't part of data Amazon examines when drawing maps.'

"historical racial divide"

Atlanta



Boston



"primarily black neighborhood"

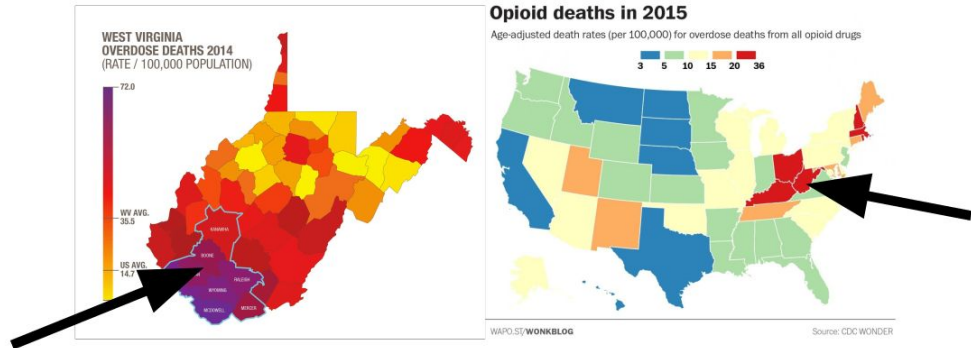
Lack of Standards (Duty of Care)

Industry-specific

"Any system or network that has been optimized for advertisements has been implicitly optimized for spreading misinformation."

Drug firms drop 20.8M pain pills on 2,900 Americans

- Pharma industry carefully tracks distribution of pills
- 20 million pills to 2 pharmacies of tiny town **obviously harmful**
- Knew harms and continued dropping. **They knew**



“The state has the highest drug overdose death rate in the nation. More than 880 people fatally overdosed in West Virginia in 2016.”

Lack of Standards (Duty of Care)

Local and Regional

1984: removed limits to harmful children advertising, said adversaries must not be judged for content that targets receive

1988: vetoed overwhelming support for limits to harmful advertising, said any protection of children from harm would oppress rights of advertisers

- Congress: must address “ideological child abuse” risk
- Broadcasters: “we expected the President to sign it”
- POTUS: attackers must not be subject to “tastes of agency officials”



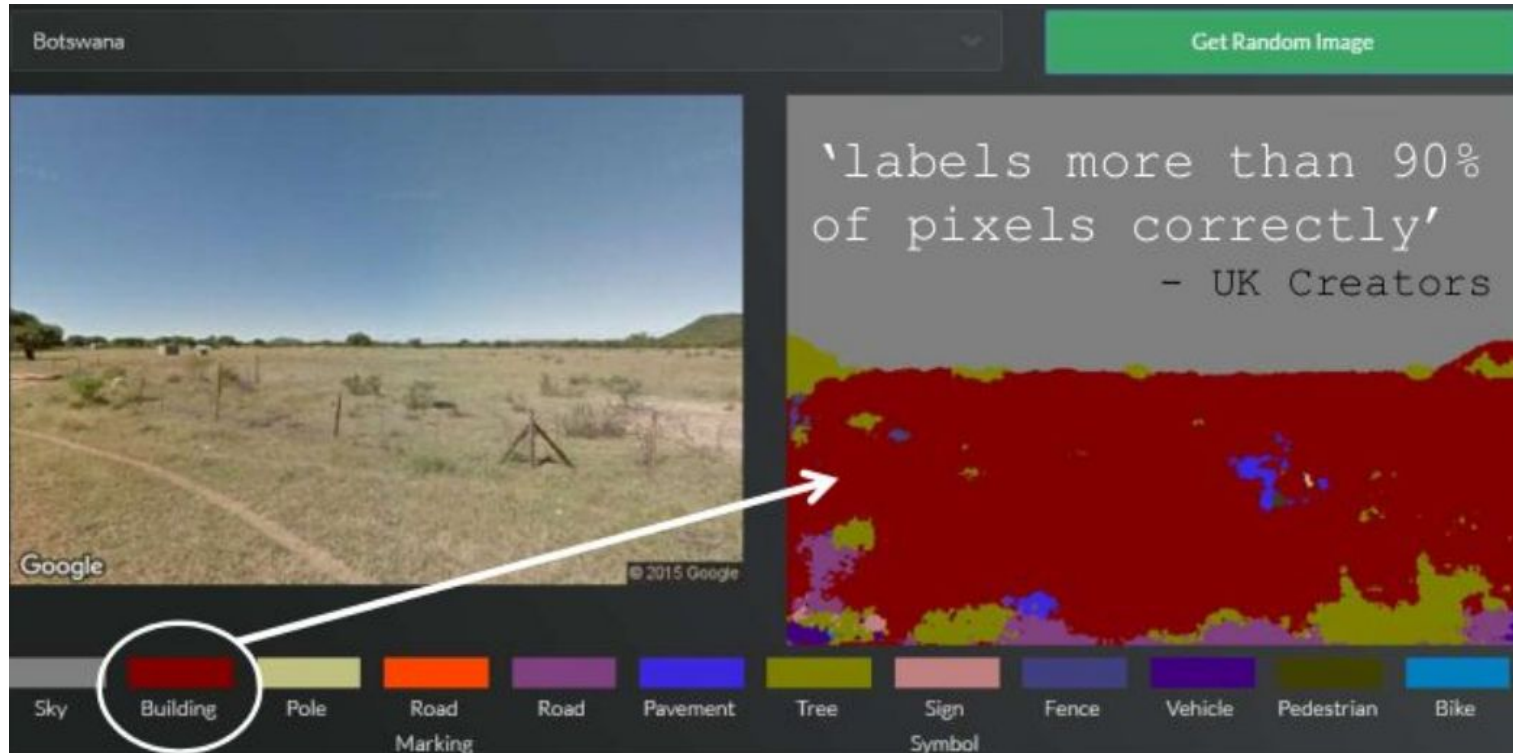
Overconfidence (Taybot Had a Backdoor)

It's
“Learning”

11:32 PM - 23 Mar 2016



Overconfidence



Overconfidence

Cars as Ants

“...flow control will be possible via a few mobile actuators (less than 5%) long before a majority of vehicles have autonomous capabilities...”

Using Leonia as a cut-through to the George Washington Bridge will increase your commute time. Staying on the major highways will be quickest route to the George Washington Bridge.



Small town uses low-tech solution to combat Waze

Navigation apps like Waze and Google Maps can help speed up your comm

CBSNEWS.COM

Overconfidence

 [Redacted] Apr 17
Owner video of Autopilot steering to avoid collision with a truck

 **Autopilot Saves [Redacted]**
[Redacted] autopilot eaved the car autonomously from a side collision from a boom lift truck. I was driving down the interstate and you can see the boom L...
youtube.com

← ↻ 2.4K ❤️ 5.8K ...

 (((davi - 德海))) @daviottenheimer - Apr 17
@cwhite_92 @lindsayceil [Redacted] they're behaving in predicate manner shifting toward exit..why watch until late block them and then freak?

← ↻ ❤️ 📺 ...

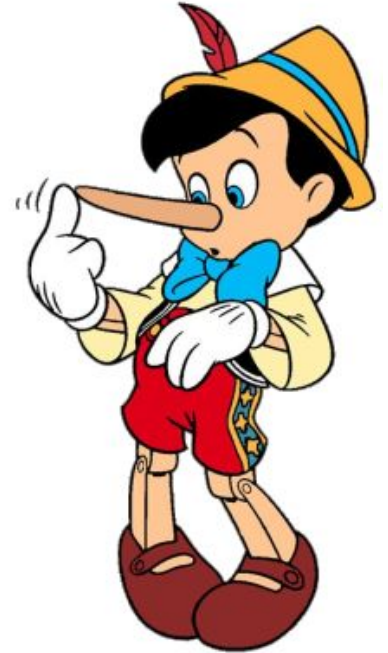
 (((davi - 德海)))
@daviottenheimer

@tjdonegan @cwhite_92 @lindsayceil
[Redacted] story i see is proper analytics (eg human) far earlier detects what [Redacted] blind to until late

6:39 PM - 17 Apr 2016

Overconfidence (Repeatedly Lying to Congress)

“In April, shortly before Zuckerberg’s Senate testimony, Phandeeyar and five other Myanmar groups **blasted him for claiming** in an interview with Vox that Facebook’s **systems had detected and removed incendiary messages** in September last year. ‘We believe **your system, in this case, was us,**’ they wrote. Zuckerberg apologized.”



Conclusion: Here's *Where* to Hold the Audit Line

Automation Systems (AI and/or ML)

1. Must have transparency to preserve human freedom and rights
2. Must come with human training for emergency response procedures including all failure modes, alarms and controls
3. Must be assessed for behavior risks, beyond transactional/functional security, for humans to assess “entirety” of complex likelihood/severity
4. Must respect boundaries, never be single points of failure (must not hide cascading unknown functions that cause escalating harms)
5. Given 1-4, safety is cultural and must always override economy

Conclusion: Here's *Why* to Hold the Audit Line

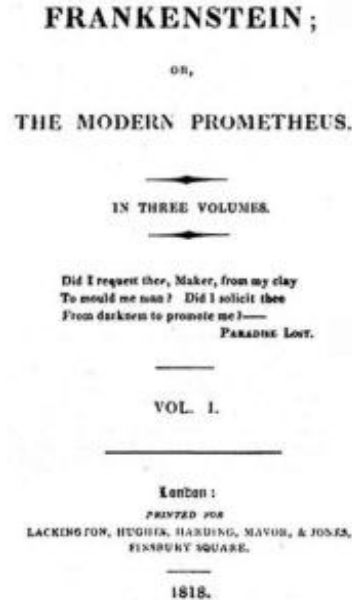
1816: “humans might not use incredible technological advances responsibly”



2007: “...the brave, as yet unnamed officer was unable to stop the wildly swinging computerised Swiss/German Oerlikon 35mm MK5 anti-aircraft twin-barrelled gun.

By the time the gun had emptied its twin 250-round auto-loader magazines, nine soldiers were dead and 11 injured.

The unknown officer tried to shut the gun down but she couldn't because the computer gremlin had taken over.”



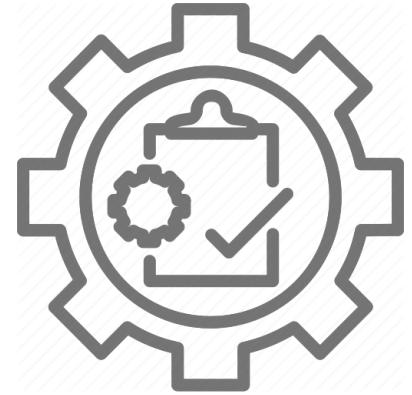
A Few Tools for Tests

Organization-based

- IBM: <https://developer.ibm.com/open/projects/ai-explainability/>
- Google: <https://pair-code.github.io/what-if-tool/>
- DotEveryone: <https://doteveryone.org.uk/project/consequence-scanning/>
- U Chicago: <https://dsapp.uchicago.edu/projects/aequitas/>

Github

- <https://github.com/marcotcr/lime>
- <https://github.com/adebayoj/fairml>
- <https://github.com/slundberg/shap>
- <https://github.com/LASER-UMASS/Themis>
- <https://github.com/Thenerdstation/mltest>
- <https://github.com/suriyadeepan/torchtest>
- <https://github.com/tensorflow/cleverhans>
- <https://github.com/shromonag/FalsifyNN>



Auditing AI:

The Civil Rights Battle of Our Time

Davi Ottenheimer