mindthesec

2021

mindthesec
2021

Why Did The Driverless Car Cross The Road?

Davi Ottenheimer

mindthesec 2021

Why Did The Driverless Car Cross The Road?

To Crash on the Other Side

Davi Ottenheimer

A Dangerous Game of Chicken

(quick review of ethics for crossing the road)

Aristotle

To actualize its potential

Hume

Out of custom and habit

Sartre

True to self it acted in good faith

Wittgenstein

Crossing encoded into the objects “chicken” and “road”

Abstract

Looking at past examples, learning from failures, is meant to ensure that we avoid their repetition.

Lately we must ask if even our best machines need serious corrective help from humans in order to learn from their own past. What is the threshold for repeatedly failing a basic rule before we accept these products may never improve their safety?

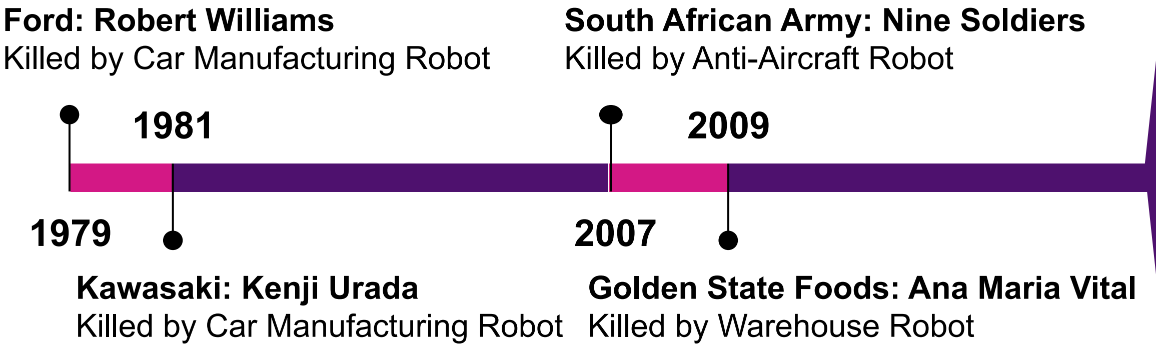
It turns out (as should be expected) if someone focuses complex machines narrowly while claiming general abilities, and then ignores safety decision controls or similar values, they will simply accelerate avoidable disasters and people die faster.

At the end of the presentation you may find yourself wondering how many lives security teams could have saved by applying basic ethics to machine development when it claims to be “learning”.

About me

Davi Ottenheimer is Vice President of Trust and Digital Ethics at Inrupt. Prior to Inrupt Davi led development of client-side field-level encryption for a non-relational database. He has been engaged in delivering safety within the cultural disruptions of emerging technology for over 25 years as the head of security, managing global security engineering, operations and assessments, as well as over a decade of incident response and digital forensics. Davi received his postgraduate academic Master of Science degree in International History from London School of Economics, focused on ethics of military intervention.

Real Killer Robot History



1979

Ford: Robert Williams
Killed by Car Manufacturing Robot

1981

Kawasaki: Kenji Urada
Killed by Car Manufacturing Robot

2007

South African Army: Nine Soldiers
Killed by Anti-Aircraft Robot

2009

Golden State Foods: Ana Maria Vital
Killed by Warehouse Robot

2015

- **VW: Anonymous**
  Killed by Car Manufacturing Robot
- **SKH Metals: Ramji Lal**
  Killed by Welding Robot

2016

- **Dallas Police: Micah Johnson**
  Killed by Bomb Defuser Robot
- **Ajin USA: Regina Elsea**
  Killed by Car Manufacturing Robot
- **Tesla: Gao Yaning**
  Killed by Autopilot Car
- **Tesla: Joshua Brown**
  Killed by Autopilot Car

2017+

- **VIM: Wanda Holbrook 2017**
  Killed by Car Manufacturing Robot
- **Tesla: Yoshihiro Umeda 2018**
  Killed by Autopilot Car
- **Uber: Elaine Herzberg 2018**
  Killed by Autopilot Car
- **Boeing: 346 Passengers 2019**
  Killed by Autopilot Plane

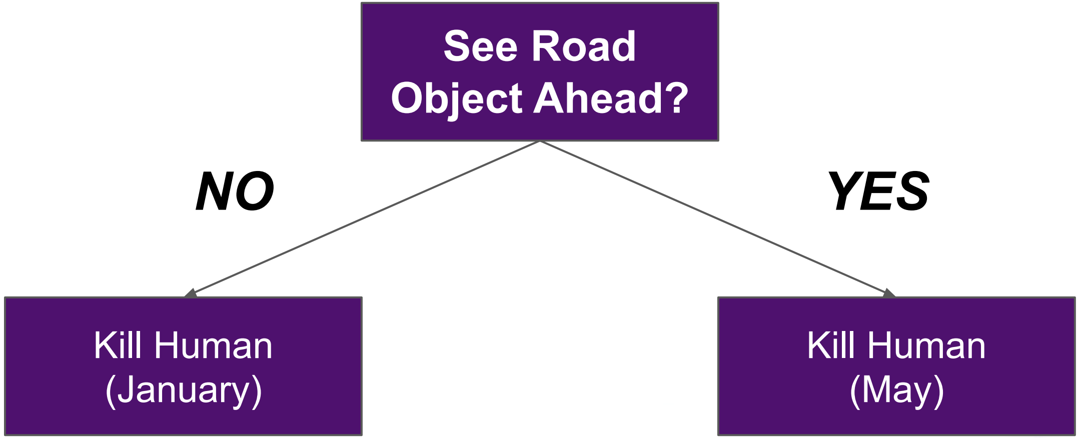
Alleged Tesla “Autopilot” Deaths Total: 20

(as of August 2021)

https://www.tesladeaths.com/
https://www.eetimes.com/are-we-prepared-for-killer-robo-drivers/

HOW?

2016: “Autopilot” Safety Decision Tree



See Road Object Ahead?

NO

YES

Kill Human (January)

Kill Human (May)

High visibility service vehicle with flashing safety lights

法治封面
“自动驾驶”：安全，不安全！？
9月14日
星期三
事故如何发生 记录仪显示行车轨迹
12:38
车流量比较大的几条高速分别为：京港澳高

https://www.flyingpenguin.com/?p=34838

“I don’t know why he went over to the slow lane...”

https://www.flyingpenguin.com/?p=22441

2016: “Autopilot” Predictions *After* Two Deaths

June: “I really would consider autonomous driving to be basically a solved problem. I think we're basically ***less than two years away from complete autonomy***.”

October: “Full autonomy will enable a Tesla to be substantially ***safer than a human driver***… We are excited to announce that, as of today ***all Tesla vehicles… have the hardware needed for full self-driving*** capability at a safety level substantially greater than that of a human driver.”

https://www.youtube.com/watch?v=wsixsRI-Sz4&t=4675s
https://www.tesla.com/blog/all-tesla-cars-being-produced-now-have-full-self-driving-hardware

Shooting the Safety Messengers

2016: “Writing an article that’s negative, you’re effectively dissuading people from using autonomous vehicles, you’re killing people.”

2018: “It's really incredibly irresponsible of any journalists with integrity to write an article that would lead people to believe that autonomy is less safe because people might actually turn it off, and then die.”

2021: “Elon Musk says Tesla’s latest beta self-driving software is ‘not great’...”

https://www.wired.com/story/tesla-autopilot-safety-statistics/
https://www.cnbc.com/2021/08/23/elon-musk-says-tesla-fsd-beta-9point2-software-is-not-great.html

False & Unsafe Propaganda

- September 2014: “They will be a factor of 10 safer than a person [at the wheel] in a six-year time frame.”
- December 2015: “We’re going to end up with complete autonomy, and I think we will have ***complete autonomy in approximately two years***.”
- June 2016: “I really consider autonomous driving a solved problem, I think we are less than two years away from complete autonomy, safer than humans.”
- March 2017: “I think that [you will be able to fall asleep in a Tesla] in about two years.”

https://www.flyingpenguin.com/?p=33693

2021: “Autopilot” Reality

“Tesla has advertised its cars since 2016 as already having all the hardware necessary for all the FSD features to be used in the future without any hardware upgrades. In reality, ***only cars delivered since the spring, 2019*** have the necessary hardware…”

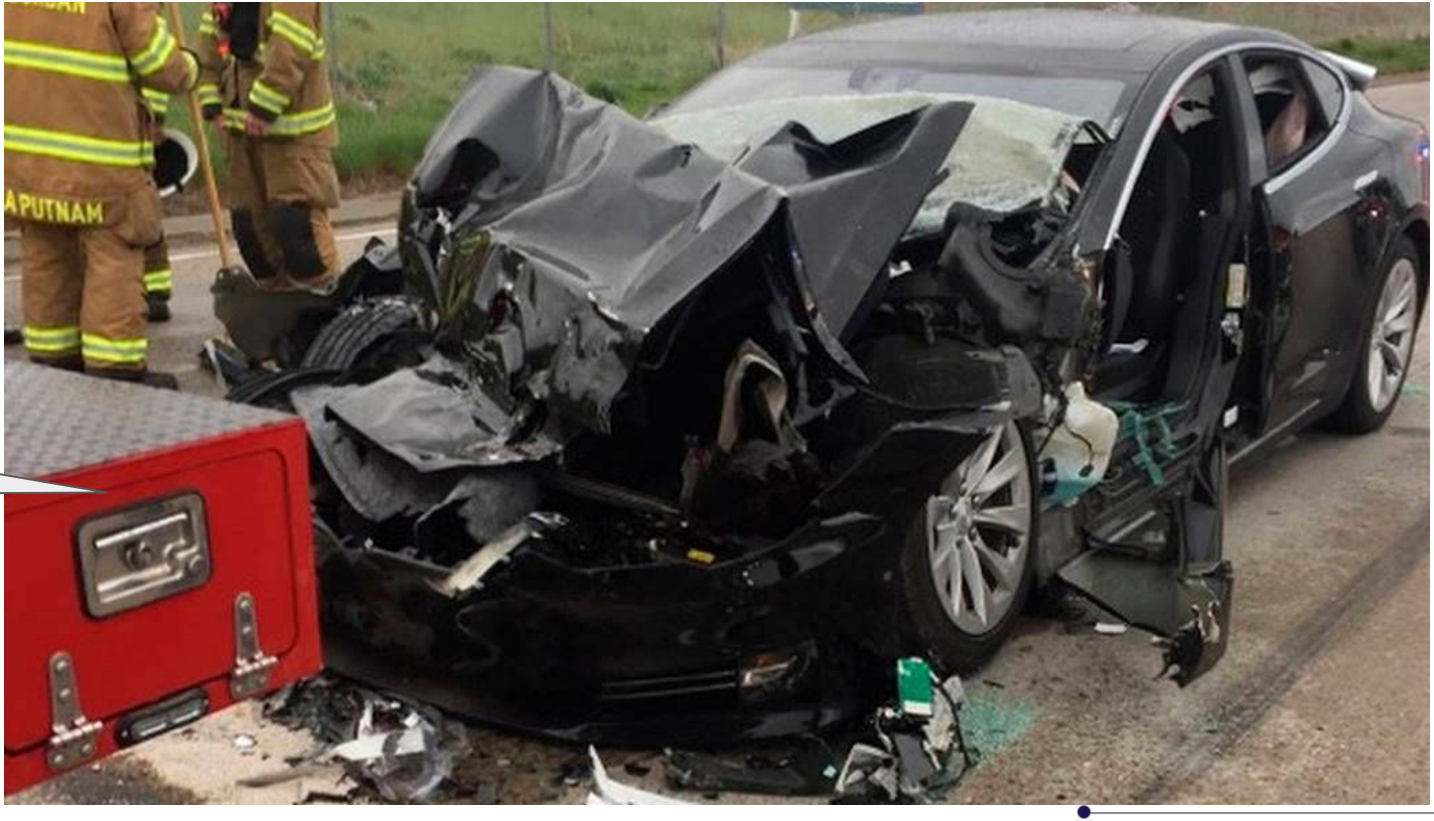
“Videos of Tesla drivers testing the system show ***the system struggles with basic driving tasks…***”

“If something goes wrong with Autopilot, it's ***because*** someone is misusing it…” -- Elon Musk

https://www.autoweek.com/news/green-cars/a37059597/tesla-fsd-subscribers-might-need-dollar1500-in-new-hardware/
https://www.businessinsider.com/tesla-senators-ftc-self-driving-autopilot-marketing-claims-elon-musk-2021-8

EXAMPLE INVESTIGATIONS OF “NO”

2018: “Autopilot” Drives Into Back of Fire Truck



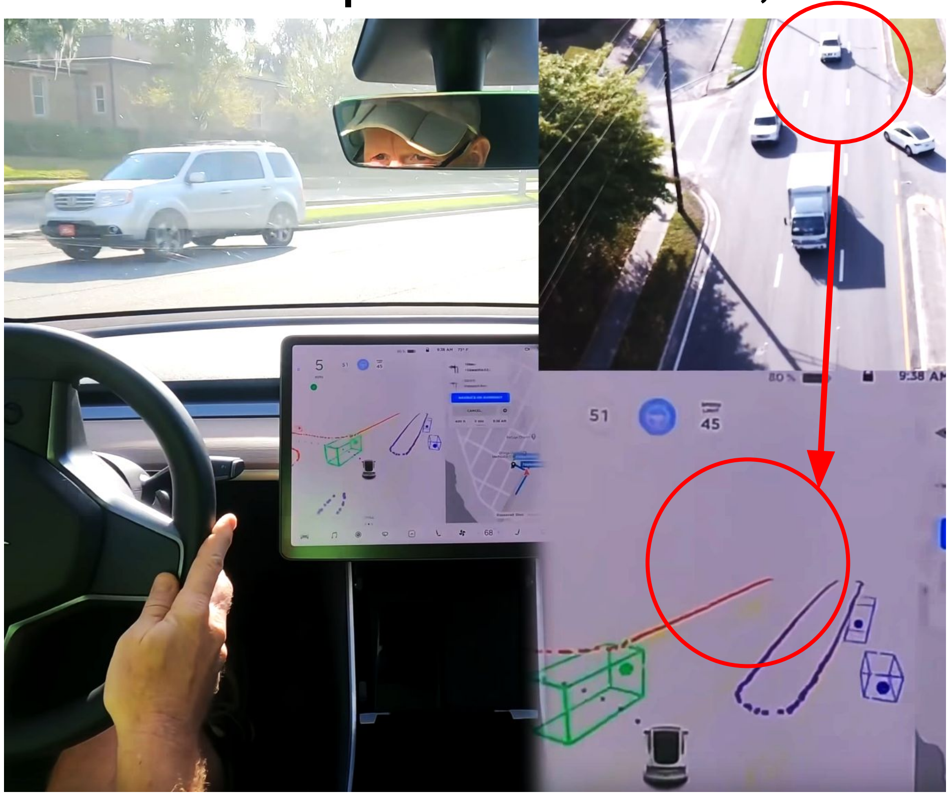
High visibility service vehicle stopped at red light

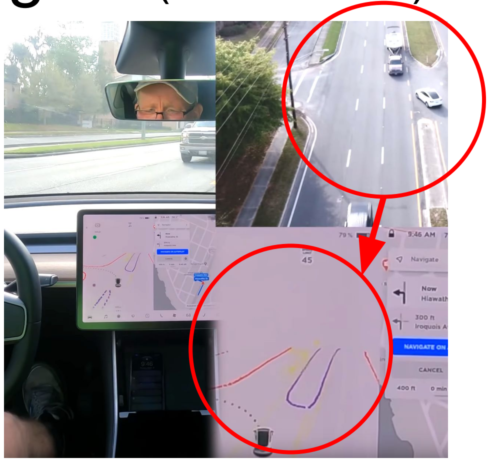
2021: NHTSA Documents 11 “NO” Cases In 3 Years

Eleven Tesla vehicles made between 2014 and 2021 “encountered first responder scenes and subsequently struck one or more vehicles. [...] Crash scenes encountered [since 2018] included scene control measures such as first responder vehicle lights, flares, an illuminated arrow board, and road cones.”

https://static.nhtsa.gov/odi/inv/2021/INOA-PE21020-1893.PDF

2021: “Autopilot” So Blind, Is It Legal? (March 18th)





Human Override: About to Crash Into Oncoming Traffic

“Full speed… going straight for that… STOP! Aaaargghh!”

https://youtu.be/uClWlVCwHsI?t=243
https://youtu.be/uClWlVCwHsI?t=508

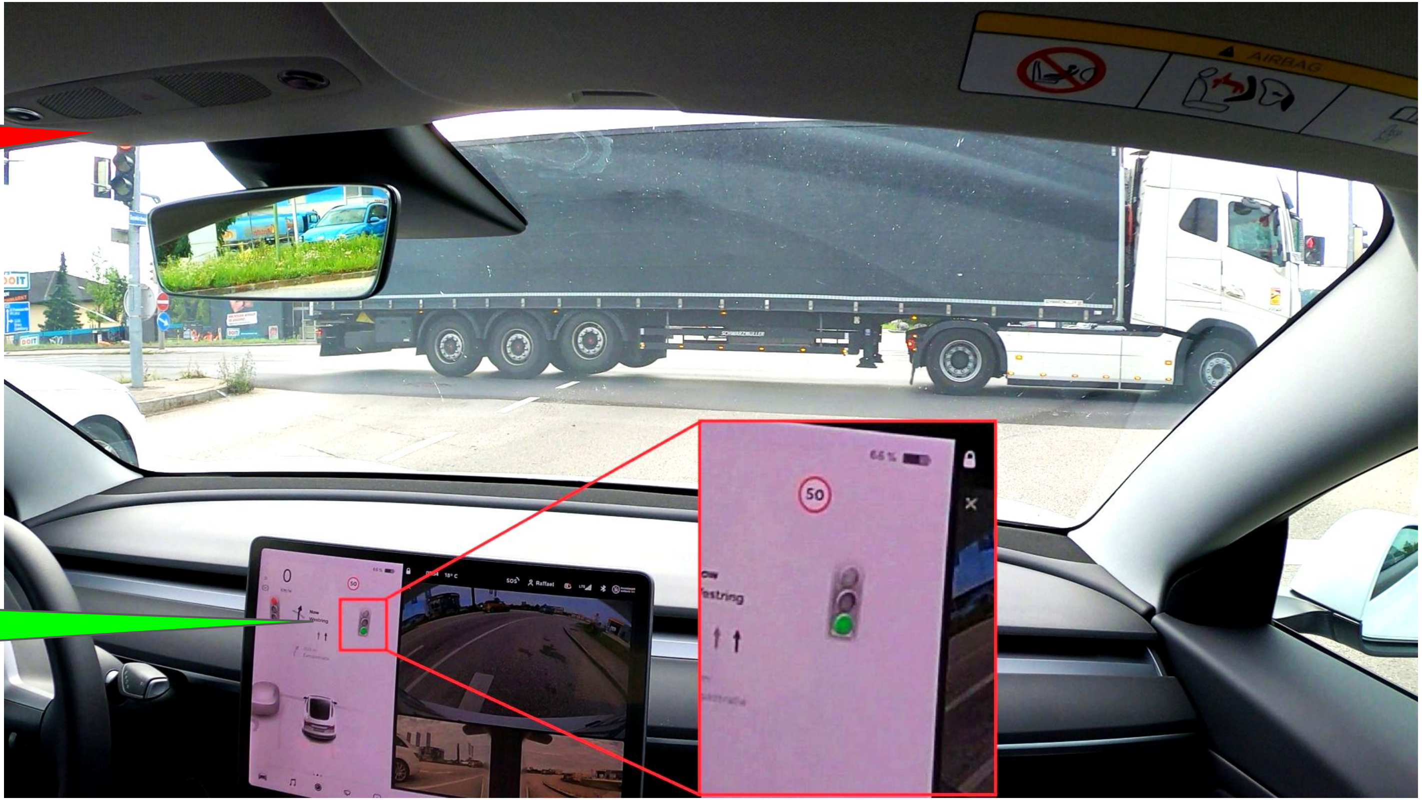
2021: “Autopilot” So Blind, Is It Legal? (April 18th)

- 0.92 seconds (~100ft): computer recognizes light change to yellow
- 1.88 seconds (~200ft): computer applies brakes to slow
- 2.00 seconds “Kick Out” warning...

Computer disables itself at 47 mph, speeding towards red light intersection

https://www.flyingpenguin.com/?p=33894

2021: “Autopilot” So Blind, Is It Legal? (August 4th)



Red Light

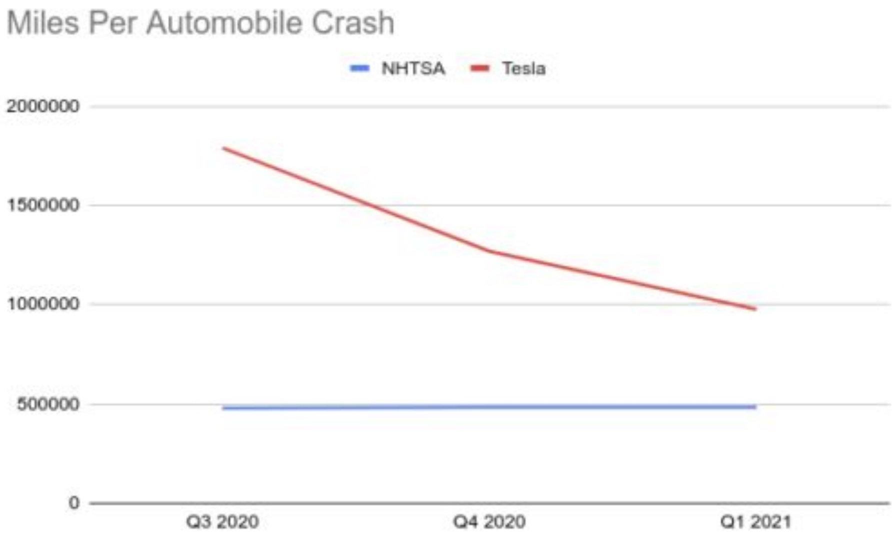
Green Light

https://www.flyingpenguin.com/?p=33894

2021: “Autopilot” So Blind, Is It Even *Learning*?

“…we haven’t done too much continuous learning. We train the system once, fine tune it a few times and that sort of goes into the car. We ***need something stable*** that we can evaluate extensively and then we think that that is good and that goes into cars. So ***we don’t do too much learning*** on the spot or continuous learning.”

-- Tesla AI Day, August 19th, 2021



https://youtu.be/tSa1kOOELrY?t=2764
https://www.flyingpenguin.com/?p=33067

EXAMPLE INVESTIGATIONS OF “YES”

2021: Different Hardware, Different Software...

Same Fatal Crashes Repeatedly *Trying to Navigate* Under Trailers

2021: March 11th… 3am white trailer with safety markings

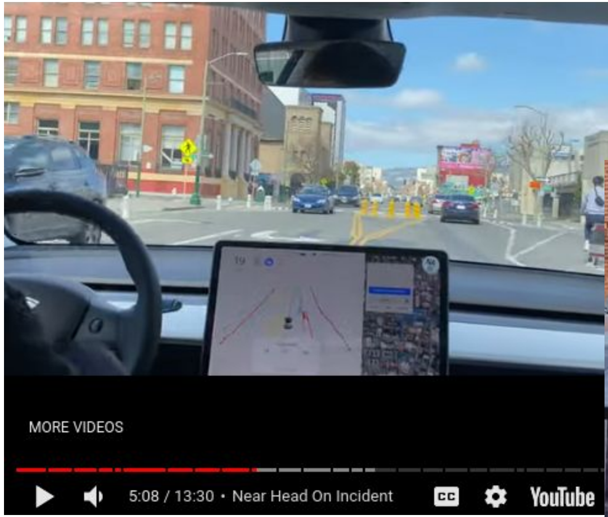
2019: March 1st (Jeremy Banner)

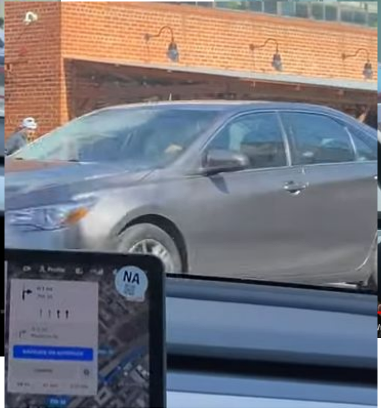
2016: May 7th (Joshua Brown)

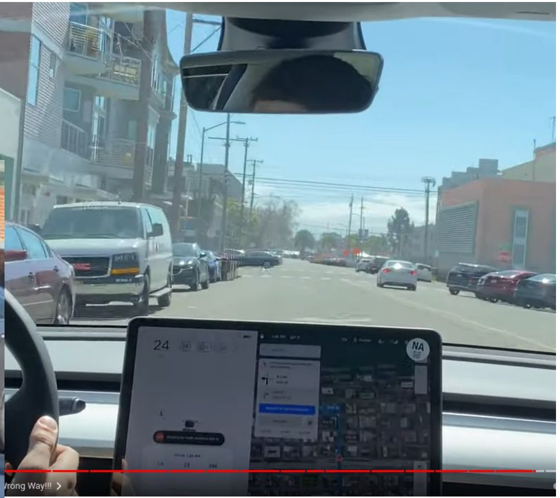
2021: “Autopilot” Decisions So Unsafe, Is It Learning?

1) Crosses double-yellow towards oncoming traffic

2) Drives wrong side of road







3) “Holy Shhh*” two near collisions

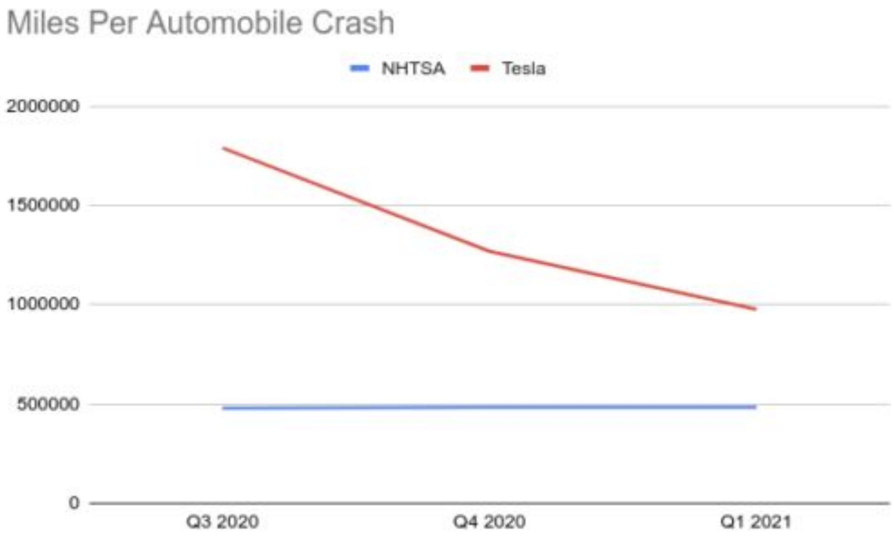
https://youtu.be/antLneVlxcs

WHY?

Beware The “We Think That That Is Good”

“…we haven’t done too much continuous learning. We train the system once, fine tune it a few times and that sort of goes into the car. We need something stable that we can evaluate extensively and then ***we think that that is good and that goes into cars***. So we don’t do too much learning on the spot or continuous learning.”

-- Tesla AI Day, August 19th, 2021



https://youtu.be/tSa1kOOELrY?t=2764
https://www.flyingpenguin.com/?p=33067

Inherited Versus Controlled Rights

Rights Within Inherited System

Authority Can be Judged, Found Wrong and Held Accountable

Rights Within Controlled System

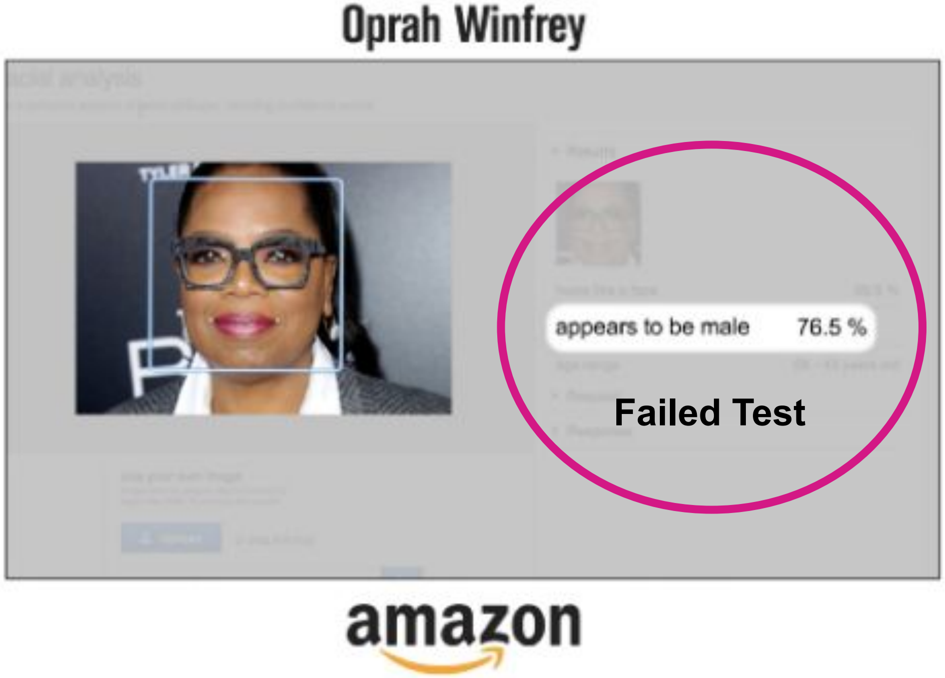
Authority is the Judge and Can *Never be Wrong* or Held Accountable

“...we think that that is good…”

Example of "Controlled System" Speech

"The answer to anxieties over new technology is ***not to run 'tests'*** inconsistent with how the service is designed to be used…"

-- Matthew Wood, AWS AI General Manager



appears to be male 76.5 %

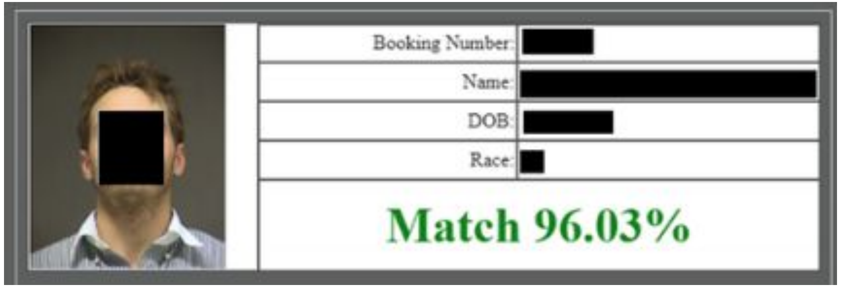
https://aws.amazon.com/blogs/aws/thoughts-on-machine-learning-accuracy/?ascsubtag=f20aee8a83dfe170e3c0efee8cdf84898de11b97
https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced

Example of “Controlled System” Speech

“Amazon has ***repeatedly claimed that the researchers failed*** to use the software, called Rekognition, ***in the way the company has instructed police to use it*** [with a 99% confidence threshold].

However, the only law enforcement agency Amazon has acknowledged as a client says it also does not use Rekognition in the way Amazon claims it recommends...”

Washington County Sheriff’s Office in Oregon Public Information Officer: “We do not set nor do we utilize a confidence threshold.”



Booking Number: [redacted]

Name: [redacted]

DOB: [redacted]

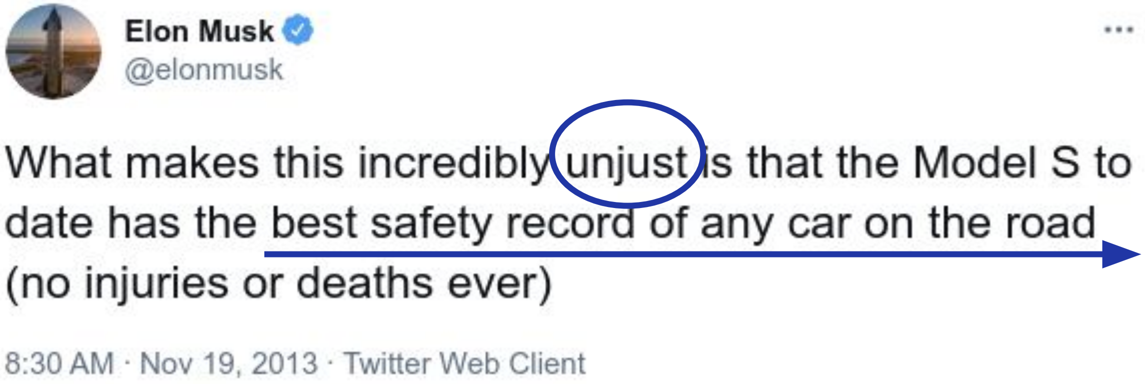
Race: [redacted]

Match 96.03%

“Amazon spokesperson clarified that ***law enforcement clients failure*** to use a 99-percent confidence threshold ***does not constitute an irresponsible application***...”

https://gizmodo.com/defense-of-amazons-face-recognition-tool-undermined-by-1832238149

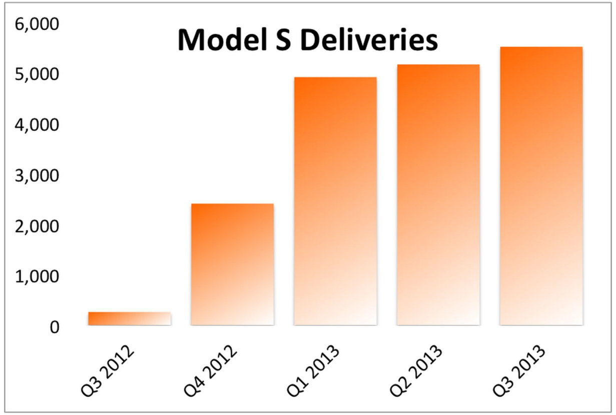
Example of “Controlled System” Speech



Elon Musk
@elonmusk

What makes this incredibly unjust is that the Model S to date has the best safety record of any car on the road (no injuries or deaths ever)

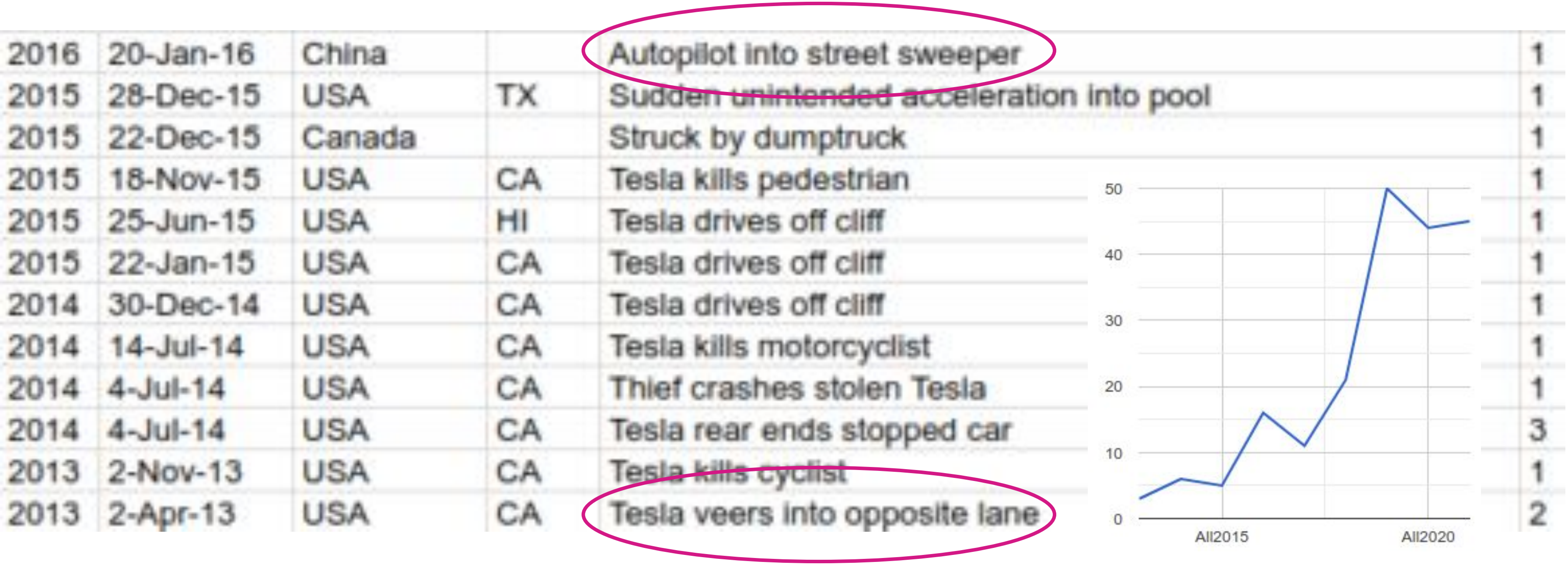
8:30 AM · Nov 19, 2013 · Twitter Web Client



- 2013 Model S veers into opposite lane killing 2
- 2014 Big claims about “Autopilot” having (unverified) “safety” features
  - Tesla “Model S” recorded ***over 25 deaths*** as of March 2021
  - Other brands same period more perfect safety records than ever before

https://www.forbes.com/sites/afontevecchia/2013/11/12/elon-musk-on-the-tesla-fires-headlines-are-deceiving-model-s-is-safest-car-on-the-road-by-far/
https://www.iihs.org/ratings/vehicle/tesla/model-s-4-door-hatchback/2021

Example of “Inherited System” Speech



2016	20-Jan-16	China		Autopilot into street sweeper	1
2015	28-Dec-15	USA	TX	Sudden unintended acceleration into pool	1
2015	22-Dec-15	Canada		Struck by dumptruck	1
2015	18-Nov-15	USA	CA	Tesla kills pedestrian	1
2015	25-Jun-15	USA	HI	Tesla drives off cliff	1
2015	22-Jan-15	USA	CA	Tesla drives off cliff	1
2014	30-Dec-14	USA	CA	Tesla drives off cliff	1
2014	14-Jul-14	USA	CA	Tesla kills motorcyclist	1
2014	4-Jul-14	USA	CA	Thief crashes stolen Tesla	1
2014	4-Jul-14	USA	CA	Tesla rear ends stopped car	3
2013	2-Nov-13	USA	CA	Tesla kills cyclist	1
2013	2-Apr-13	USA	CA	Tesla veers into opposite lane	2

As soon as Tesla was on the road it had to start reporting injury and death

https://www.tesladeaths.com/

Example of “Inherited System” Speech

- **Tesla ‘Autopilot’ leads to *more* crashes than regular driving**
- Tesla Model S has higher insurance losses than other large luxury cars” (higher frequency and severity)
- Tesla has fire deaths at 4x the rate of other vehicles
- Teslas have 2-4x more non-crash fires than the average car, and incur damages up to 7x higher
- Teslas have 3x driver deaths of comparably priced luxury vehicles
- Tesla crashes twice as often as regular cars
- Tesla Deaths as of 7/7/2021: 201 (Verified Autopilot: 9)

https://www.tesladeaths.com/

Trivial Inexpensive Attacks Defeat Best Cameras

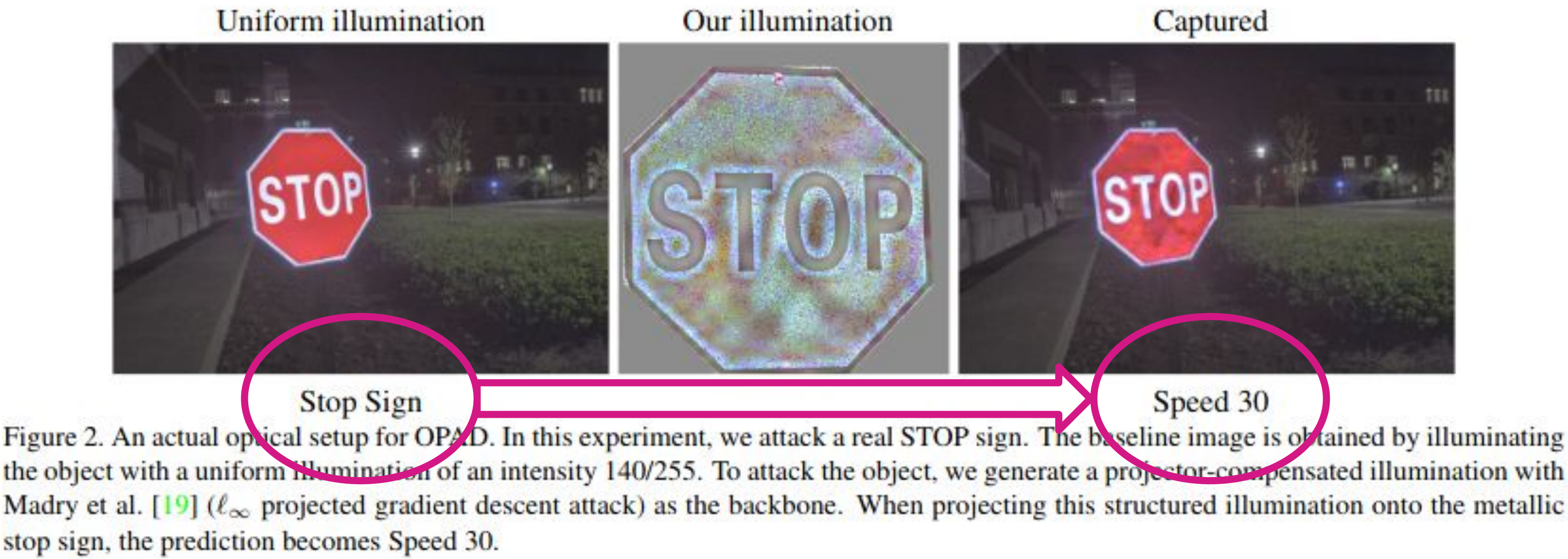
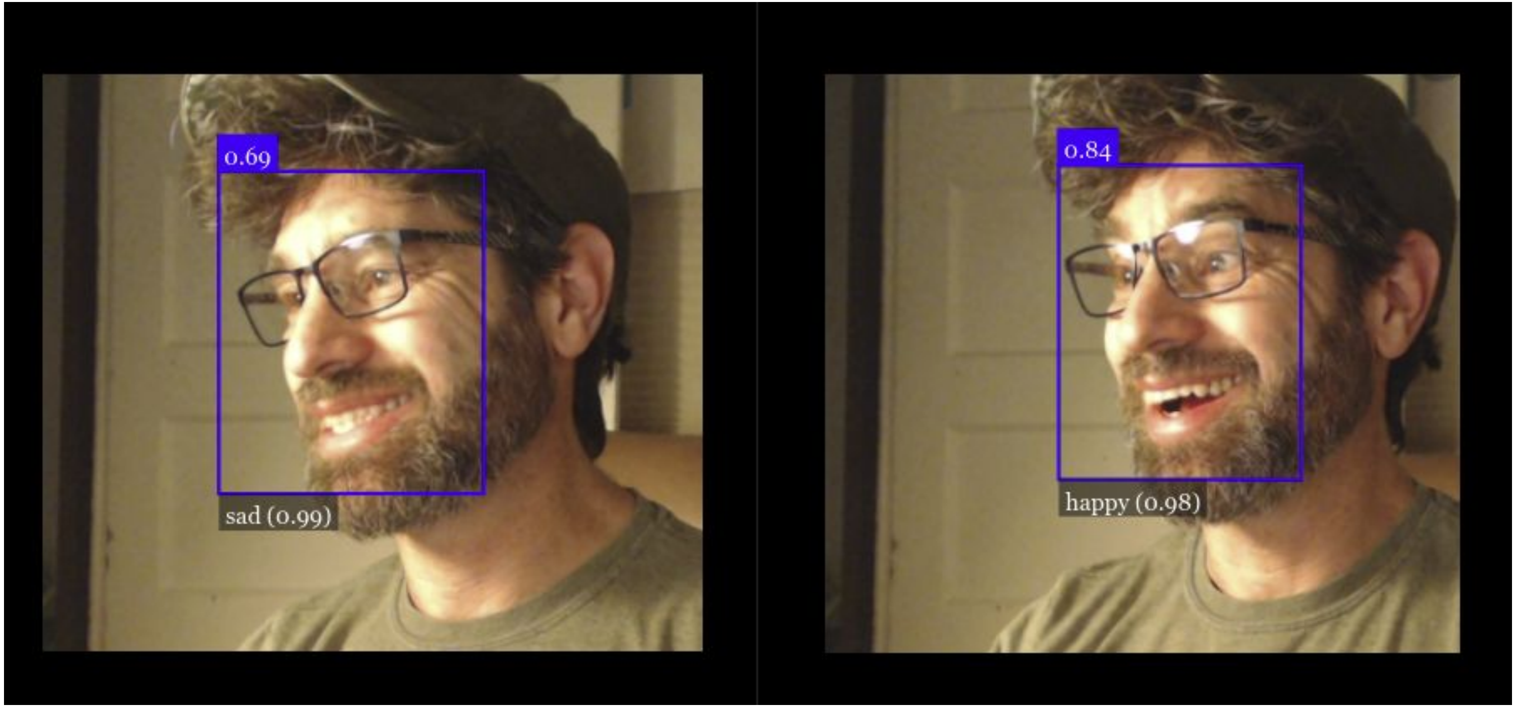


Figure 2. An actual optical setup for OPAD. In this experiment, we attack a real STOP sign. The baseline image is obtained by illuminating the object with a uniform illumination of an intensity 140/255. To attack the object, we generate a projector-compensated illumination with Madry et al. [19] (ℓ_∞ projected gradient descent attack) as the backbone. When projecting this structured illumination onto the metallic stop sign, the prediction becomes Speed 30.

https://arxiv.org/pdf/2108.06247.pdf

Trivial Inexpensive Attacks Defeat Best Cameras



Neither image is true...

https://emojify.info/menu
https://www.flyingpenguin.com/?p=22441

“Bug Bounty” ***Sugar Coating Flaws*** of Algorithms (Bias)

“Kulynych, the winner of the prize, said he had mixed feelings about the competition. ‘Algorithmic harms are not only ‘bugs’. Crucially, a lot of harmful tech is harmful not because of accidents, unintended mistakes, but rather by design. This comes from maximisation of engagement and, in general, profit externalising the costs to others. As an example, amplifying gentrification, driving down wages, spreading clickbait and misinformation are not necessarily due to ‘biased’ algorithms.’”

https://www.theguardian.com/technology/2021/aug/10/twitters-image-cropping-algorithm-prefers-younger-slimmer-faces-with-lighter-skin-analysis

X-Ray Algorithm Reveals Race -- **No Known Fixes**

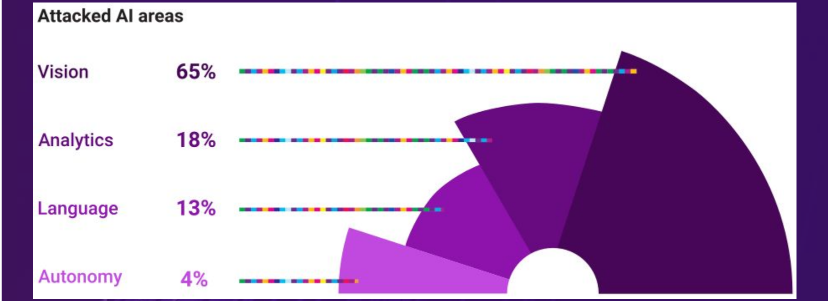
“AI systems trained to analyze X-rays, CT scans, mammograms, and other medical images were able to predict a patient’s self-reported race with a high degree of accuracy based on the images alone… even when the images they were analyzing were degraded to the point that anatomical features were indistinguishable to the human eye.

Most concerningly, according to the paper’s authors, the team was unable to explain how the AI systems were making their accurate predictions.”

https://www.vice.com/en/article/wx5ypb/ai-can-guess-your-race-based-on-x-rays-and-researchers-dont-know-how

Four Answers as to Why...

1. **Easily Corrupted Systems (Controlled Rights)**
2. **Trivial Inexpensive Attacks**
3. **Sugar Coating Algorithmic Flaws as "Bugs"**
4. **No Known Fixes**



Source: https://incidentdatabase.ai/summaries/incidents
https://adversa.ai/report-secure-and-trusted-ai

mindthesec 2021

Thank you!

✉ davi

in daviottenheimer

◎ what is this?

f are you kidding me? no.