- CAN A.I. THINK ETHICALLY?

By 2050, driverless cars and mobility-as-a-service will grow to $7 trillion worldwide

THE DRIVERLESS DILEMMA

FROM 2035 TO 2045

Consumers will regain up to 250 MILLION HOURS OF FREE TIME from behind the wheel

$234 BILLION IN PUBLIC COSTS will be saved by reducing accidents from human error

Driverless cars can ELIMINATE 90% OF TRAFFIC FATALITIES — Saving 1 million lives every year

HOW WILL DRIVERLESS CARS DETERMINE WHOSE LIFE SHOULD BE SPARED?

MORALITY IN MACHINES

Driverless cars 'must decide quickly, with incomplete information, in situations that programmers often will not have considered, using ethics that must be encoded ahead of time.'

NOAH J. GOODALL
Senior Research at the Virginia Transportation Research Council

Who should A.I. save?

IN A GLOBAL STUDY, MOST PEOPLE PREFERRED

- Swerving over staying the course
- Sparing passengers over pedestrians
- Saving as many lives as possible

Participants were most likely to spare the lives of a child, and least likely to spare animals and criminals

MINIMIZED HARM VS. PASSENGER PROTECTION

76%

of people felt driverless cars should SAVE AS MANY LIVES AS POSSIBLE

BUT, very few were willing to buy a vehicle programmed to minimize harm

They prefer cars programmed to PROTECT PASSENGERS AT ALL COSTS

DRIVERLESS CARS WILL SAVE LIVES, BUT PROGRAMMING THEM TO DO SO COULD SLOW THEIR ADOPTION AND COST MANY MORE LIVES

REAL LIFE APPLICATIONS GROW EVEN MORE COMPLEX

In an accident causing injuries but not fatalities, should A.I.

- DISTRIBUTE INJURIES EVENLY, harming more people less severely?
- CONSIDER THE LIKELIHOOD and severity of potential injuries?
- Take into account the QUALITY OF LIFE EFFECT of resulting injuries?
- Is it better to HOSPITALIZE 5 PEOPLE, OR KILL 1?

AS A.I. ADVANCES, IT BECOMES RESPONSIBLE FOR MORE MORAL AND ETHICAL DECISION MAKING

WHEN A.I. GOES WRONG

AMAZON'S REKOGNITION

REKOGNITION'S FACIAL RECOGNITION ALGORITHMS CAN

- Identify up to 100 faces in a single image
- Track people in real time through surveillance cameras
- Scan footage from police body cameras

IN 2018, THE ACLU COMPARED 25,000 MUG SHOTS TO PHOTOS OF EVERY MEMBER OF CONGRESS USING REKOGNITION - THEY FOUND

28 false matches

39% Were people of color, who make up just 20% of Congress

LAW ENFORCEMENT AGENCIES ARE CALLED AVIDLY USING REKOGNITION

- Orlando Police Department (Florida)
- Washington County Sheriff's Office (Oregon)
- In 2016, half of Americans adults were included in a law enforcement facial recognition network

WHAT IF YOUR FACE WASN'T REALLY YOU?

STAR WARS: IN 2016, ROGUE ONE RECREATED ACTORS FROM THE ORIGINAL FILM USING CGI

- GRAND MOFF TARKIN: Originally portrayed by Peter Cushing, died in 1994
- PRINCESS LEIA ORGANA: Originally portrayed by Carrie Fisher, at age 19

Most critics felt the recreations WEREN'T QUITE BELIEVABLE

DEEP FAKES

- In January 2018, YouTube DEEPFAKES recreated these scenes from Rogue One using a standard desktop computer — achieving a better result than the original

- In April 2018, BuzzFeed created a fake video of BARACK OBAMA saying "We're entering an era in which our enemies can make it look like anyone is saying anything at any point in time."

- In May 2018, a Belgian political party released a fake video of DONALD TRUMP commenting on Belgium's climate policy

Using the "deepfakes" tech, anyone can create fake videos of real people

So far, these videos have been voiced by impersonators, but, more convincing fakes are coming

GOOGLE ASSISTANT

In May 2018
GOOGLE DEBUTED DUPLEX, A NEW FEATURE OF GOOGLE ASSISTANT

- Makes outgoing calls to schedule appointments and reservations
- Convincingly imitates human speech with 'um' and 'mmhmm'

At the same time,
GOOGLE ANNOUNCED JOHN LEGEND WOULD BE THE NEW VOICE OF GOOGLE ASSISTANT

- Google captured Legend's voice during short recording sessions
- A.I. built a digital voice model, capable of saying anything

WITH A.I., YOUR FACE - AND VOICE - COULD BE USED TO CREATE A FAKE VIDEO, IDENTIFIED USING A FACIAL RECOGNITION ALGORITHM, AND USED TO FALSELY CONVICT YOU OF A CRIME

How do we know what's real?

BUILDING TRUSTWORTHY TECH

HOW A.I. BIAS HAPPENS

WOMEN NEED APPLY

Starting in 2014, Amazon began training an A.I. to review job candidates

- The system was trained using resumes submitted over 10 years — MOST CAME FROM MEN
- The A.I. concluded that 'male' was a preferred quality for new hires, and started FILTERING OUT FEMALE CANDIDATES

DETECTING DARKER SKIN TONES

In a 2018 study, researchers found that the object detection models used in driverless cars were better at identifying pedestrians with lighter skin

- The study used a standardized set of photos to train their A.I. — but found their DATASET CONTAINED 3X AS MANY LIGHT SKINNED PEOPLE
- The A.I. quickly learned to identify light skinned pedestrians, but STRUGGLED TO IDENTIFY DARKER SKIN TONES

BOTH RECRUITING AND PEDESTRIAN DETECTION ALGORITHMS FAILED BECAUSE THEY WERE TRAINED ON BAD DATA — A.I. LEARNED BIAS FROM HUMANS

MAKING ETHICAL A.I.

START WITH DATA

A.I. training data must reflect real diversity and control for existing bias

- Amazon's recruiting algorithm was trained to eliminate female candidates. INSTEAD, IT COULD HAVE BEEN PROGRAMMED TO IGNORE GENDER
- Pedestrian identification algorithms struggle to identify darker skin tones. Rather than maximizing success overall, the TRAINING DATA COULD HAVE WEIGHTED DARK SKIN DATA POINTS MORE HEAVILY

CONSIDER THE PROCESS

When training A.I., programmers typically split their dataset into 2 parts

- Half is used to TRAIN THE A.I.
- Half is used to VERIFY AND MEASURE SUCCESS

If the initial dataset is flawed, the test will have the same bias

MONITOR FOR UNKNOWNS

Programmers must monitor for unintentional bias appearing in their A.I.

- Subtle patterns can lead A.I. to PERPETUATE HUMAN BIAS
- Amazon's recruiting algorithm preferred VERBS LIKE "EXECUTED" AND "CAPTURED" — WHICH TEND TO BE MORE USED BY MALES

SOMEDAY, A.I. MAY MAKE ETHICAL DECISIONS AND FUNCTION WITH TOTAL AUTONOMY

ARE YOU READY FOR THE A.I. REVOLUTION?

SOURCES

projectsuitcase.org
arxiv.org/abs/1907.10597.pdf
arxiv.org/abs/1810.03993v2.pdf
gizmodo.com/the-ethics-of-driverless-cars-1826487
gizmodo.com/people-are-terrible-at-teaching-driverless-cars-to-kill-1826154
nature.com/articles/s41598-018-20797-9#ref17
economist.com/the-world-if/2018/10/20/perhaps-anyone
npr.org/2018/05/14/610254471/people-deploy-science-ethical-questions
theguardian.com/2018/jun/14/uncover-google-assistant-plots-legend-voice
nfl.gizmodo.com/this-video-scares-the-power-of-deepfakes-to-ex-cop-in-th-1828692062
podcast.com/2018/12/31/#1600620-hope-one-new-year-great-stuff-tedx-princeton-kids
spectrum.ieee.org/transportation/self-driving/can-you-program-ethics-into-a-self-driving-car
newsroom.ucla.edu/releases/ucla-psychologists-develop-test-of-moral-decision-making
technologyreview.com/s/612776/this-is-how-ai-bias-really-happened-and-why-its-so-hard-to-fix
arxiv.org/ftp/arxiv/papers/1810/1810.03993.pdf
medium.com/@ericbrown/amazon-ai-recruiting-algorithm-discriminates-against-women-a8d2b0c0a0a0

presented by
CYBER SECURITY DEGREES

INFOGRAPHIC BY
NOWSOURCING