

AI Weapon Systems and Our Future

Abdullah Barhoum

Normative Reasoning and Machine Ethics WS19/20

Contents

- ◊ AI and Weapon Systems
- ◊ Different from Nuclear Weapons?
- ◊ Destabilizing aspects of AI
- ◊ Conclusion

AI Weapon Systems

- ◇ “Terminator”
- ◇ Strategic Advantage

Compared to Nuclear Weapons?

Compared to Nuclear Weapons?

- ◇ AI Weapon Systems are in comparison to Nuclear Weapons:
 - ◇ Cheaper
 - ◇ More precise
 - ◇ Less kinetic firepower
 - ◇ “invisible”
 - ◇ Serve a different purpose than nuclear weapons

The Destabilizing Aspects of AI

- ◆ Artificial Intelligence on the Battlefield: Implications for Deterrence and Surprise
- ◆ By Zachary S. Davis

The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

TECHNOLOGY NEWS

OCTOBER 10, 2018 / 5:12 AM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's ([AMZN.O](#)) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Artificial intelligence (AI)

🕒 This article is more than 3 years old

Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter

Attempt to engage millennials with artificial intelligence backfires hours after launch, with TayTweets account citing Hitler and supporting Donald Trump



▲ Tay uses a combination of artificial intelligence and editorial written by a team including improvisational comedians. Photograph: Twitter

Elle Hunt

Thu 24 Mar 2016 06:41 GMT

GOOGLE

TECH

ARTIFICIAL INTELLIGENCE

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

Nearly three years after the company was called out, it hasn't gone beyond a quick workaround

By [James Vincent](#) | Jan 12, 2018, 10:35am EST



Ground truth: Soap

Nepal, 288 \$/month

Azure: food, cheese, bread, cake, sandwich

Clarifai: food, wood, cooking, delicious, healthy

Google: food, dish, cuisine, comfort food, spam

Amazon: food, confectionary, sweets, burger

Watson: food, food product, turmeric, seasoning

Tencent: food, dish, matter, fast food, nutriment



Ground truth: Soap

UK, 1890 \$/month

Azure: toilet, design, art, sink

Clarifai: people, faucet, healthcare, lavatory, wash closet

Google: product, liquid, water, fluid, bathroom accessory

Amazon: sink, indoors, bottle, sink faucet

Watson: gas tank, storage tank, toiletry, dispenser, soap dispenser

Tencent: lotion, toiletry, soap dispenser, dispenser, after sha

Joe Biden Says ‘Poor Kids’ Are Just as Bright as ‘White Kids’



Former Vice President Joseph R. Biden Jr., whose record on race has been under scrutiny, appeared to make a gaffe by saying, “poor kids are just as bright and just as talented as white kids,” during a town hall hosted by the Asian & Latino Coalition.

Erin Schaff/The New York Times

By **Matt Stevens**

Aug. 9, 2019



The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership



■ classified as turtle ■ classified as rifle
■ classified as other



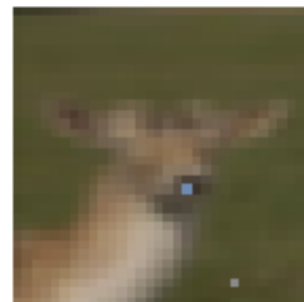
SHIP
CAR(99.7%)



HORSE
FROG(99.9%)



DEER
AIRPLANE(85.3%)



DEER
DOG(86.4%)



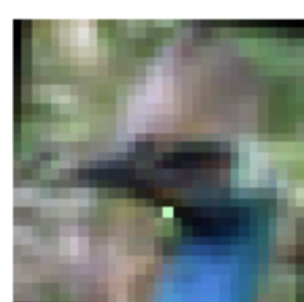
HORSE
DOG(70.7%)



DOG
CAT(75.5%)



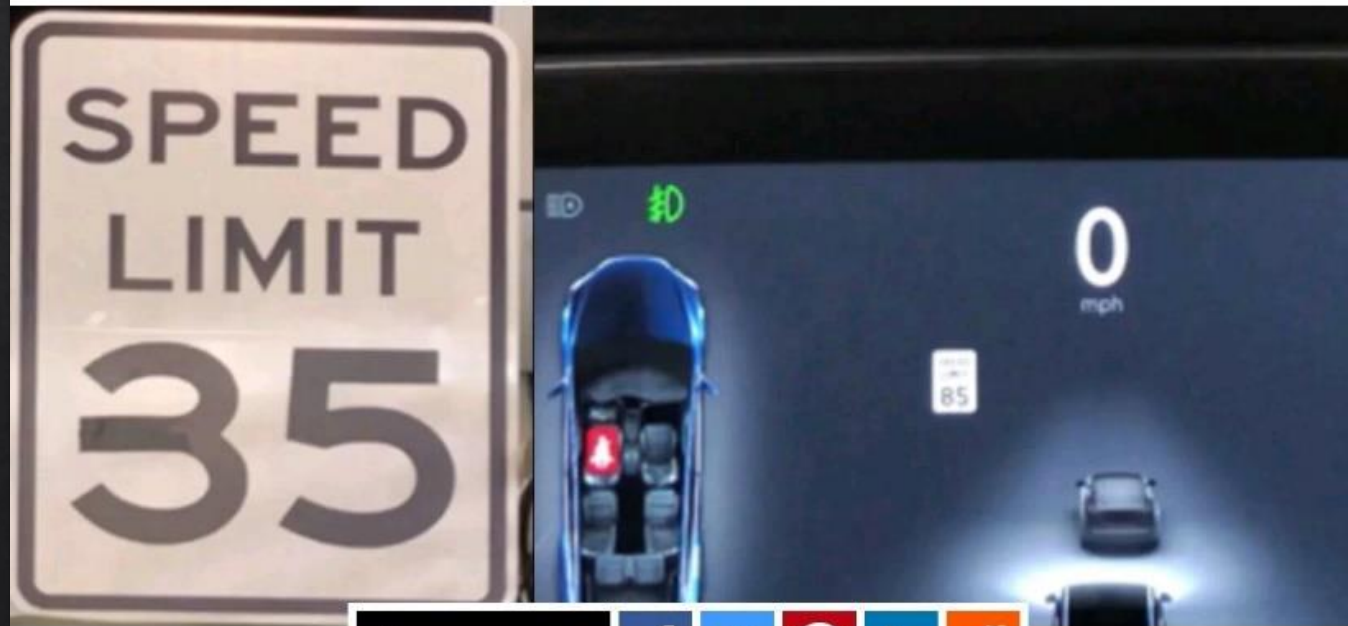
BIRD
FROG(86.5%)



BIRD
FROG(88.8%)

Tesla Autopilot gets tricked into accelerating from 35 to 85 mph with modified speed limit sign

Fred Lambert - Feb. 19th 2020 12:08 pm ET 🐦 @FredericLambert



Comments



The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

The Destabilizing Aspects of AI

- ◊ Competition will bring Uncertainty
- ◊ Data is Fragile
- ◊ Countering AI Applications can be straightforward
- ◊ Faster is not always better
- ◊ Large Scale Complexity
- ◊ AI Unpredictability
- ◊ Human in the Loop
- ◊ Public-Private Partnership

The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

Report: Palantir took over Project Maven, the military AI program too unethical for Google



by **TRISTAN GREENE** — 10 weeks ago in **ARTIFICIAL INTELLIGENCE**



<https://tnw.to/heyMV>

The Destabilizing Aspects of AI

- ◆ Competition will bring Uncertainty
- ◆ Data is Fragile
- ◆ Countering AI Applications can be straightforward
- ◆ Faster is not always better
- ◆ Large Scale Complexity
- ◆ AI Unpredictability
- ◆ Human in the Loop
- ◆ Public-Private Partnership

Conclusion

- ◇ AI Weapon Systems are complicated problem
 - ◇ Multi-hundred stakeholders
 - ◇ Many opinions and national interests
 - ◇ The human condition overpowers
- ◇ Some Organizations are working against it:
 - ◇ Ban Lethal Autonomous Weapons
 - ◇ The Campaign to Stop Killer Robots
- ◇ <https://youtu.be/UfIJGakA9v0?t=3501>

What we can do

- ◆ Just like Global Warming and Privacy
 - ◆ We wont see the effects until its too late
- ◆ Spread the awareness!
 - ◆ Break the barrier of fear of AI
 - ◆ Help people understand what AI is
 - ◆ And what its weaknesses are

Additional Material

- ◇ Amir Husain: "Is the World Ready for the Age of AI?"
 - ◇ https://www.youtube.com/watch?v=nI_JEqnsgc4
- ◇ Vince Madai: The Ethics of Artificial Intelligence - What developers should know
 - ◇ https://www.youtube.com/watch?v=U_3pMiStQkM
- ◇ Keep it Clean: Why Bad Data Ruins Projects and How to Fix it - Philip Winder
 - ◇ <https://www.youtube.com/watch?v=bUMmj85UiXM>
- ◇ Breaking black-box AI - Evelina Gabasova
 - ◇ <https://www.youtube.com/watch?v=CQMVOCTBYRA>