Artificial Intelligence, Crime and Policing

Or: *"Now That We're In The Future, What Does It Look Like?"*

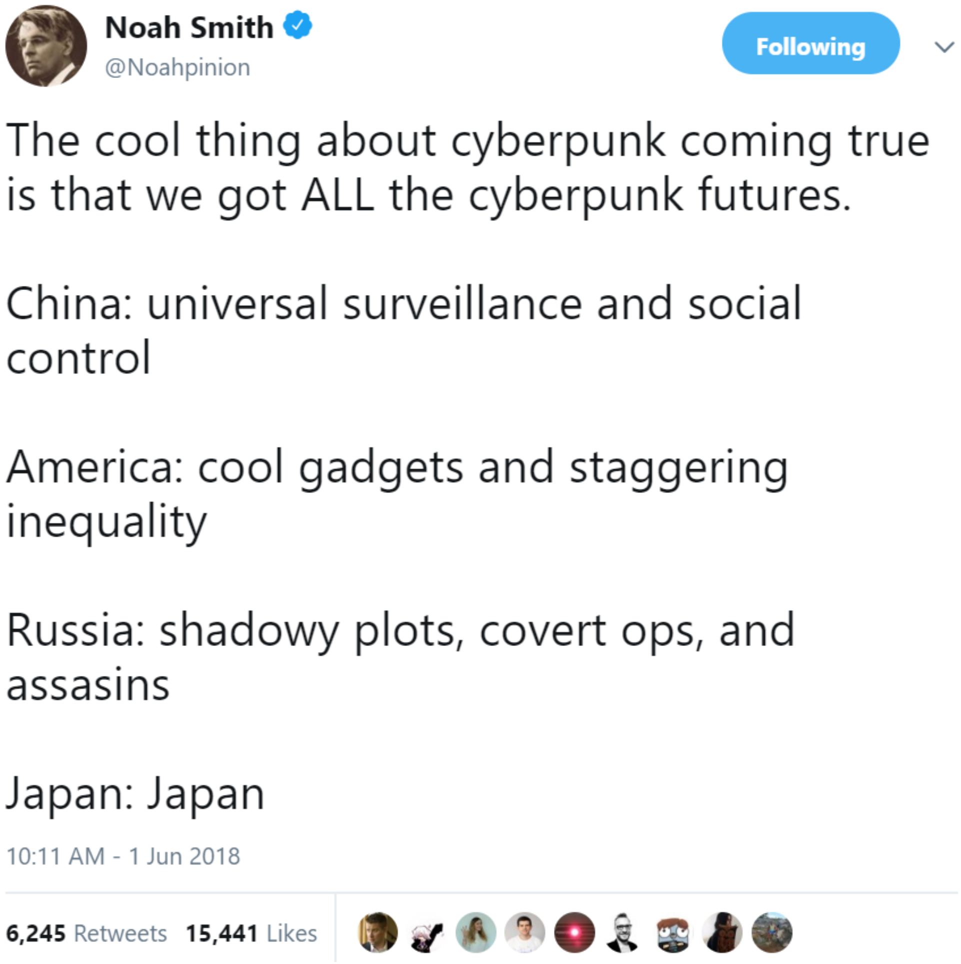
October 17th, 2018 | Copenhagen Common Sessions: Technology and Crime: Critical Criminological Perspectives

Matthijs M. Maas
<**PhD Fellow**, CILCC, Faculty of Law, University of Copenhagen>
<**Research Associate**, Governance of AI Program, Future of Humanity Institute, University of Oxford>

Matthijs.Maas@jur.ku.dk

UNIVERSITY OF COPENHAGEN

We're in the future…



Noah Smith
@Noahpinion

The cool thing about cyberpunk coming true is that we got ALL the cyberpunk futures.

China: universal surveillance and social control

America: cool gadgets and staggering inequality

Russia: shadowy plots, covert ops, and assasins

Japan: Japan

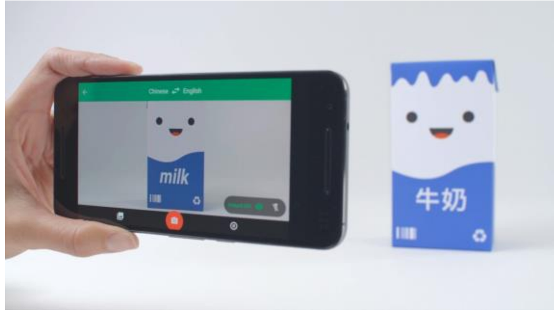
10:11 AM - 1 Jun 2018

6,245 Retweets 15,441 Likes

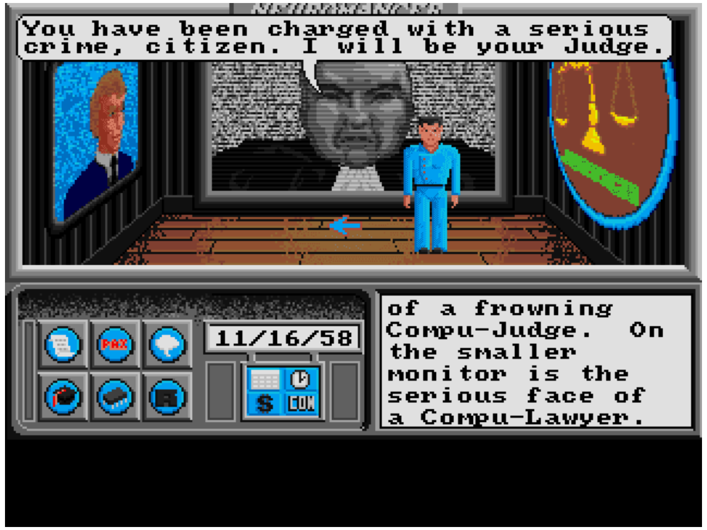
We’re in the future!

…and artificial intelligence is *everywhere*…

…and it's *confusing*!

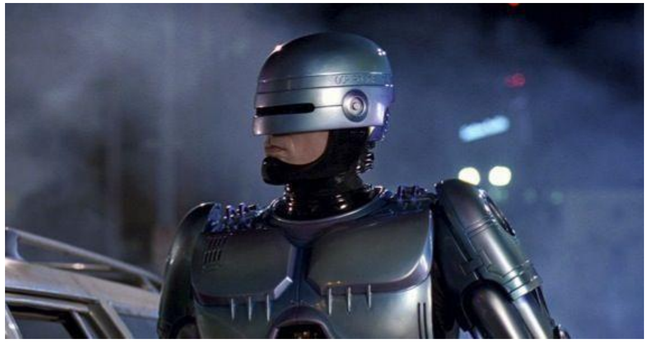


AI, crime and policing: what we were promised…


You have been charged with a serious crime, citizen. I will be your Judge.

of a frowning Compu-Judge. On the smaller monitor is the serious face of a Compu-Lawyer.

In the year 1990, old Detroit has been overtaken by crime.



…and now (that we are in the future) what does it (actually) look like?

- <goals of this talk>
  - To **reassure** you about AI—demystifying the technological dazzle, in order to better critically examine it
  - To **scare** you about AI—and its possible implications for crime/policing
  - To **provoke** discussion about the nature of 'crime', 'policing', 'fairness', and 'surveillance'

  - [Programmatic—set of examples, issues and questions for discussion]

Today

- What is AI?
- Criminal uses of AI
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

- **What is AI?**
  - Myths: what ‘intelligence’ in ‘AI’?
  - Defining AI: terms, workings
  - Uses, Requirements
- Criminal uses of AI
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

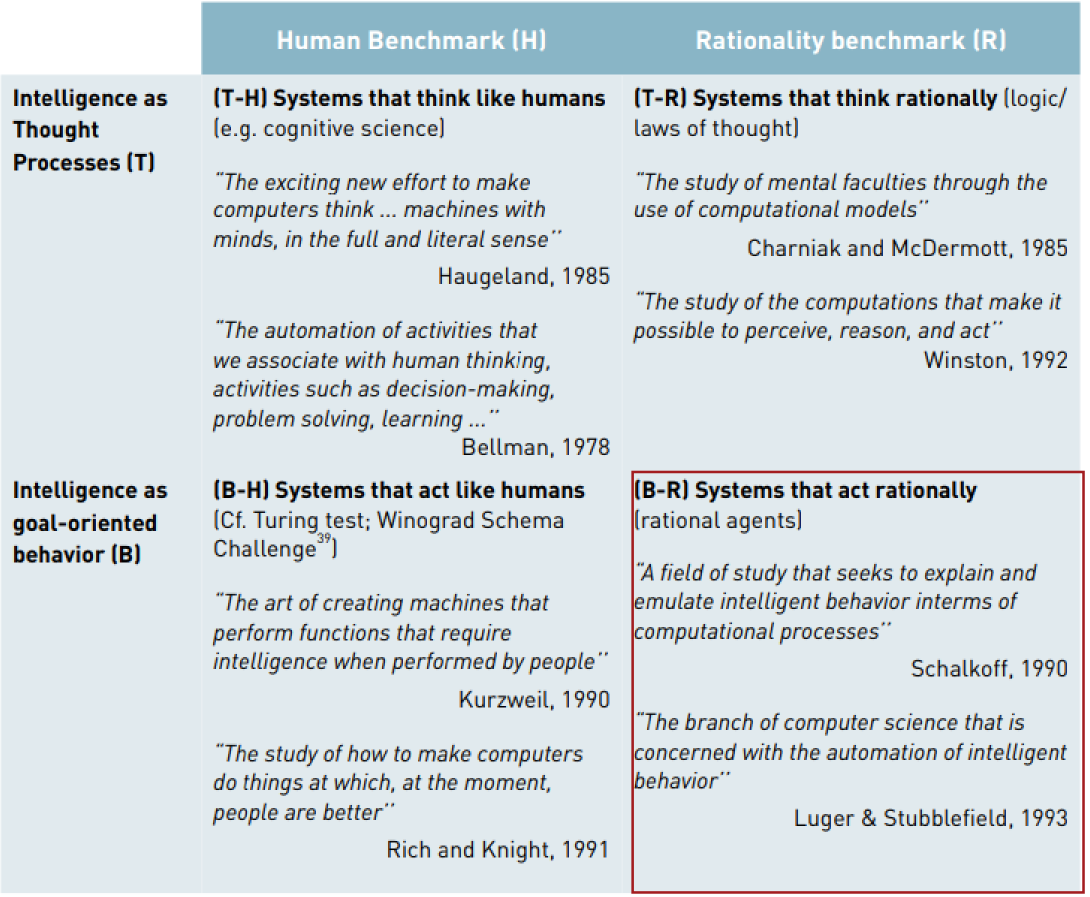
- **What is AI?**
  - **Myths: what ‘intelligence’ in ‘AI’?**
  - Defining AI: terms, workings
  - Uses, Requirements
- Criminal uses of AI
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

AI: ‘consciousness’; self-awareness’; ‘sentience’?

November 28, 2011: Qbo passes the ‘mirror self-recognition (MSR) test’, (designed to measure self-awareness in animals).

(Humanity survives)

What is Intelligence? Internal **thought process** or goal-oriented **behaviour**? **Human-** or **optimal performance?**



	Human Benchmark (H)	**Rationality benchmark (R)**
Intelligence as Thought Processes (T)	**(T-H) Systems that think like humans** (e.g. cognitive science) *"The exciting new effort to make computers think … machines with minds, in the full and literal sense"* Haugeland, 1985 *"The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning …"* Bellman, 1978	**(T-R) Systems that think rationally** (logic/ laws of thought) *"The study of mental faculties through the use of computational models"* Charniak and McDermott, 1985 *"The study of the computations that make it possible to perceive, reason, and act"* Winston, 1992
Intelligence as goal-oriented behavior (B)	**(B-H) Systems that act like humans** (Cf. Turing test; Winograd Schema Challenge[39]) *"The art of creating machines that perform functions that require intelligence when performed by people"* Kurzweil, 1990 *"The study of how to make computers do things at which, at the moment, people are better"* Rich and Knight, 1991	**(B-R) Systems that act rationally** (rational agents) *"A field of study that seeks to explain and emulate intelligent behavior interms of computational processes"* Schalkoff, 1990 *"The branch of computer science that is concerned with the automation of intelligent behavior"* Luger & Stubblefield, 1993

FIGURE 3: A TYPOLOGY OF DIFFERENT DEFINITIONS OF AI, AND THEIR UNDERLYING APPROACHES, WITH SOME EXAMPLES[39]

AI agents 'operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue [the best expected outcome]."

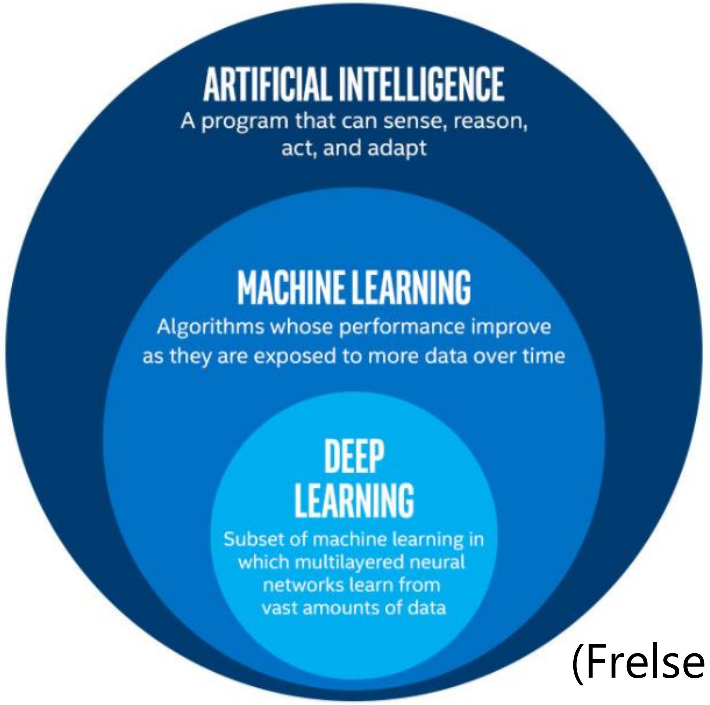
-Russell and Norvig, *Artificial Intelligence: A Modern Approach*, 4.

- **What is AI?**
  - Myths: what ‘intelligence’ in ‘AI’?
  - **Defining AI: terms, workings**
  - Uses, Requirements
- Criminal uses of AI
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

What is AI? ML? DNN?

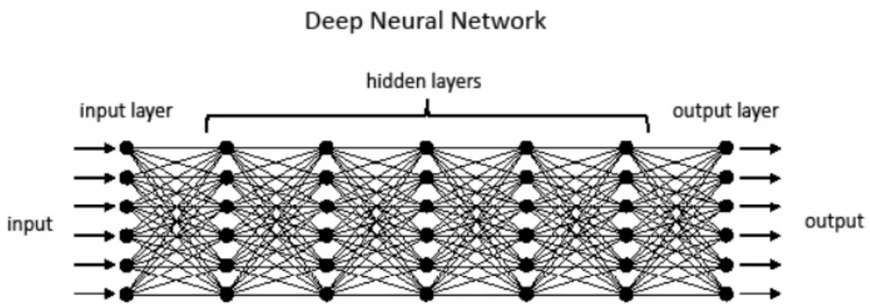
Machine learning: algorithms which adjust their behaviour (by trail and error) in reactions on training datasets, to gradually improve their performance on new data:

Deep learning is a type of machine learning, inspired by the brain, that uses multi-layer 'neural networks'...


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

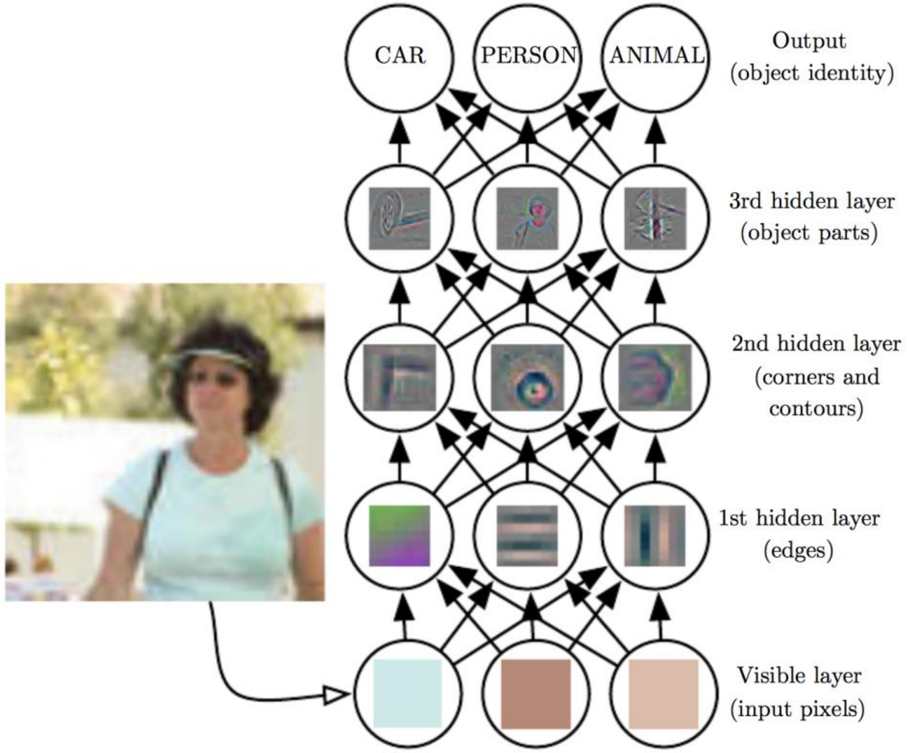
(Frelsen, 2018)

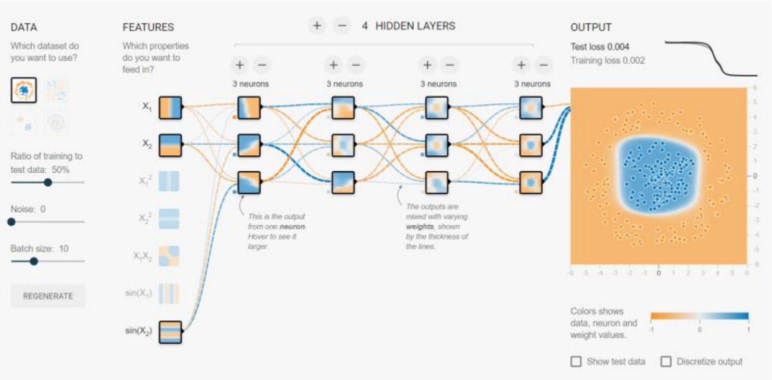
A deep learning neural network encodes knowledge, by varying the weights of connections between neurons



A deep neural network has hidden layers between the input and output layers. Some deep neural networks can have more than 150 hidden layers.

Paul Scharre, 2018





Which dataset do you want to use?

Ratio of training to test data: 50%

Noise: 0

Batch size: 10

Which properties do you want to feed in?

This is the output from one **neuron**. Hover to see it larger.

The outputs are mixed with varying **weights**, shown by the thickness of the lines.

Test loss 0.004

Training loss 0.002

Colors shows data, neuron and weight values.

□ Show test data □ Discretize output

Play around with a neural network on https://playground.tensorflow.org/

Varieties of current AI

Machine learning: algorithms which adjust their behaviour (by trail and error) in reactions on training datasets, to gradually improve their performance on new data:

- **Supervised learning**: using labelled training data [million images of 'dog', 'person', 'apple']: algorithm starts off random, but learns to distinguish relevant data features that improve its accuracy | e.g. image-recognition
- **Unsupervised learning**: machine independently identifies hidden patterns in unlabelled data, clustering them itself. | e.g. identify anomalous financial transactions; diagnose cancer in patients…
- **Reinforcement learning**: through trial-and-error in a [simulated] environment (or playing against itself), machine learns what actions give it a higher end-score… | AlphaGo (zero); Atari games
- **Generative adversarial network**: one system tries to create fake data until it is so indistinguishable from real that it fools another, trained network | DeepFakes

[there are many other approaches, including rule-based systems]

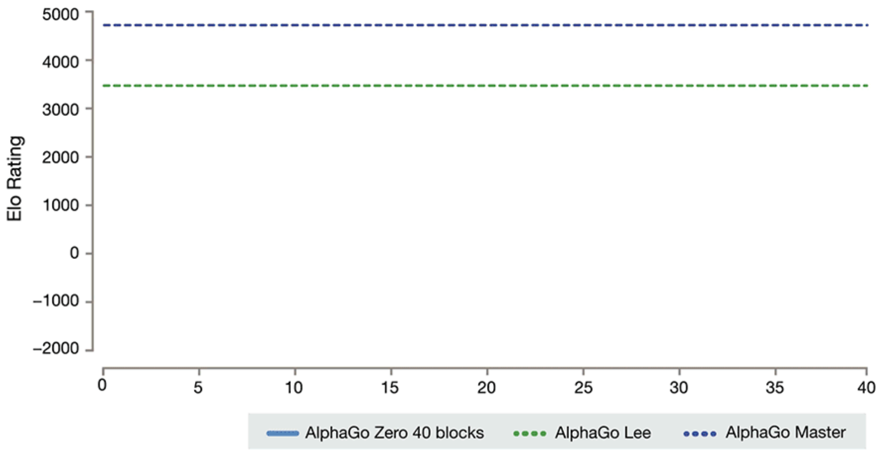
- **What is AI?**
  - Myths: what ‘intelligence’ in ‘AI’?
  - Defining AI: terms, workings
  - **Uses, Requirements**
- Criminal uses of AI
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

Uses: what AI is good for

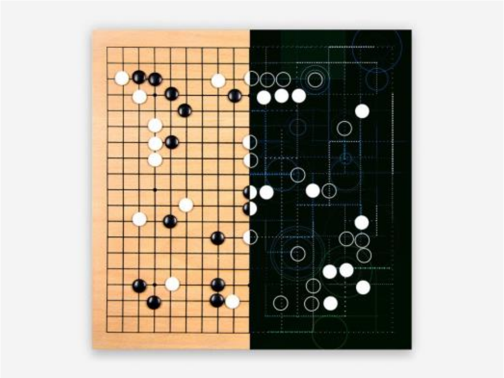
- **Data classification:** from images to spam email, and from song genres to faces
- **Anomaly detection:** fraudulent financial transactions or new malware
- **Prediction:** from Netflix and Amazon recommendations, to weather forecasting, to patient recovery rates
- **Optimization** of complex systems and tasks: e.g. energy efficiency in Google data centre
- Rapid **reaction times**, **precision** and **autonomous operation** in robots

Requirements and limits for use

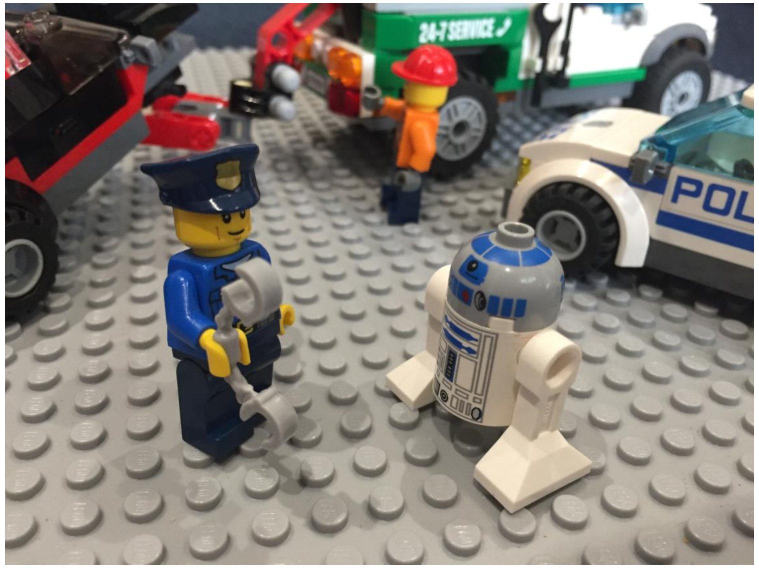
- Large and structured (or labelled) datasets
  - E.g. social media
  - E.g. financial transactions
  - E.g. re-offense rate records.
  - E.g. ability to play against itself



- Limitations:
  - Transferring learning form one task to another
  - Susceptibility to ‘adversarial input’
  - No ‘Common Sense’ → ‘artificial stupidity’



- What is AI?
- **Criminal uses of AI**
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways



- What is AI?
- **Criminal uses of AI**
  - **Crimes *with* AI: scalable cyberattacks; DeepFakes**
  - Crimes *on* AI: a new attack surface; adversarial input; ‘AI theft’
  - Crimes *by* AI: ‘algorithmic entities’ as criminal intermediary?
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

Crimes *with* AI: why will AI enable 'malicious use'? [Brundage et al. 2018]

"Expansion of existing threats.

- *The **costs of attacks may be lowered** by the scalable use of AI systems to complete tasks that would ordinarily require human labor, intelligence and expertise. A natural effect would be to **expand the set of actors** who can carry out particular attacks, the rate at which they can carry out these attacks, and the set of potential targets.*

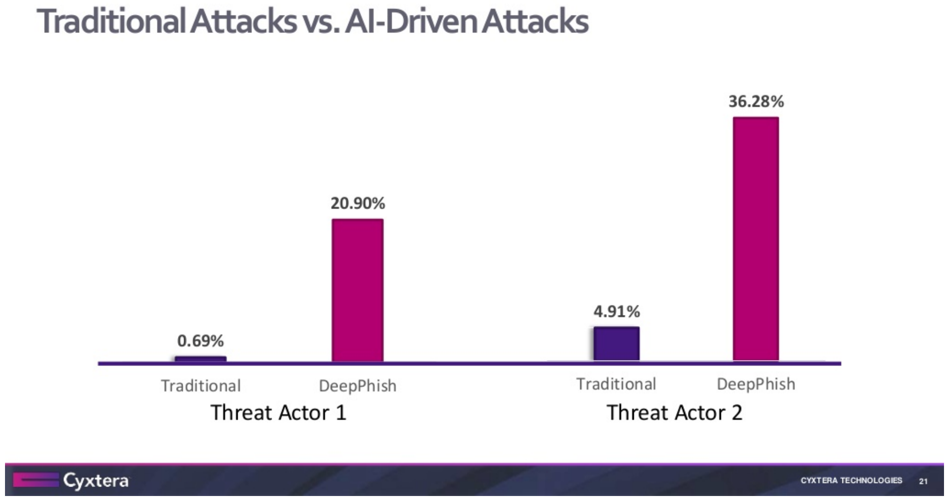
Introduction of new threats.

- *New attacks may arise through the use of AI systems to complete tasks that would be **otherwise impractical** for humans. In addition, malicious actors may **exploit the vulnerabilities of AI systems** deployed by defenders."*

'Malicious Use of AI' Report:
https://maliciousaireport.com/

Crimes *with* AI | in cyberspace: scalable hacking & identity theft

- Humans are vulnerability in any system. Today: 91% of cybercrimes and attacks start with a phishing email.
- But today, phishing emails are either generic and unconvincing (‘FBI’; ‘Nigerian Prince…’, ‘You’ve won 1$M!’), caught by spam filers, or must be (hand)tailored to targets (usually not worth time and effort).
- With ‘DeepPhish’ AI: automatically learn from, combine features (synthetic URLs, etc.) from other phishing attacks, avoiding spam filters and improving success rate…
- Eventually: automatically tailor emails to individuals?



https://albahnsen.com/wp-content/uploads/2018/05/deepphish-simulating-malicious-ai_submitted.pdf

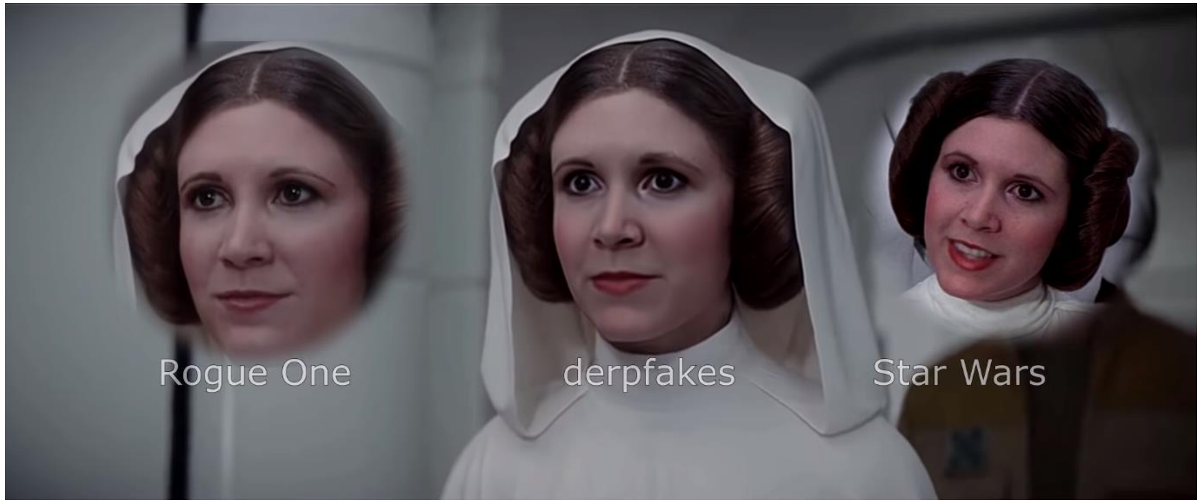
Crimes *with* AI | DeepFakes as custom blackmail material?

- Ability to edit faces into videos
- Original use: this is the internet—What do you *think*?

[···]

- Case: Leia (Carrie Fischer), inserted digitally in Star Wars *Rogue One*

- Formerly: high threshold
  - (Disney - industrial CGI)
- Now: incredibly accessible
  - (laptop)



https://www.youtube.com/watch?v=RiBqZoVe92U

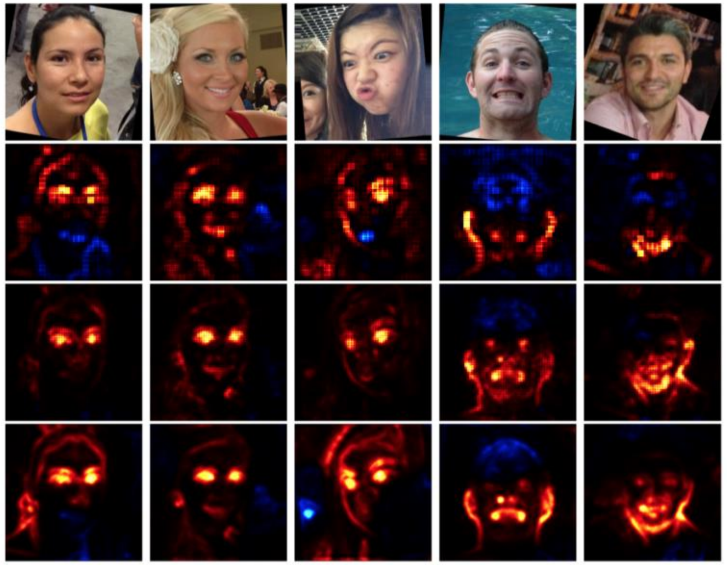
Crimes *with* AI | using AI to ‘steal’ a voice (– one day, your grandmother gets a call from you...)

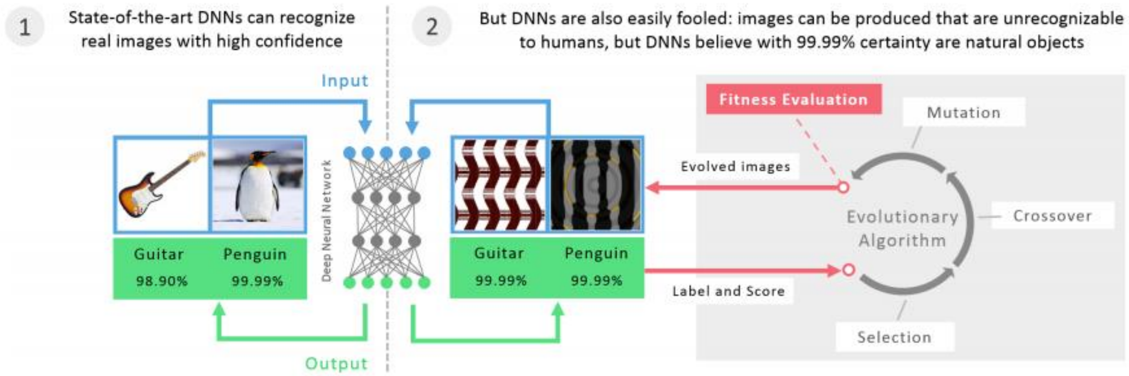
‘You won’t believe what Obama Says in this Video!’
https://www.youtube.com/watch?v=cQ54GDm1eL0

‘Clone’ your own voice (just 1 min of audio required) with *Lyrebird* at- https://lyrebird.ai/

- What is AI?
- **Criminal uses of AI**
  - Crimes *with* AI: scalable cyberattacks; DeepFakes
  - **Crimes *on* AI: a new attack surface; adversarial input; 'AI theft'**
  - Crimes *by* AI: 'algorithmic entities' as criminal intermediary?
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

Crimes *on* AI | a new attack surface for 'adversarial input' ('hypnotizing' your AI, making it 'hallucinate')





1 State-of-the-art DNNs can recognize real images with high confidence

2 But DNNs are also easily fooled: images can be produced that are unrecognizable to humans, but DNNs believe with 99.99% certainty are natural objects

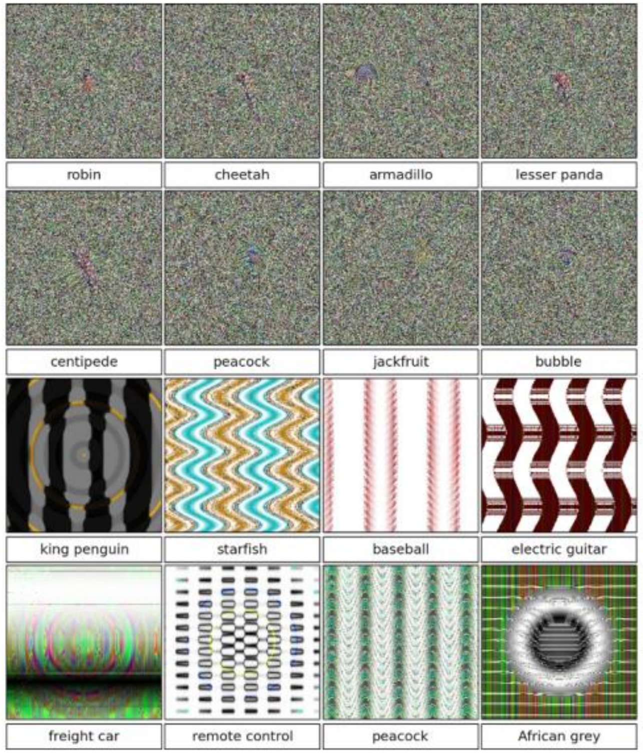
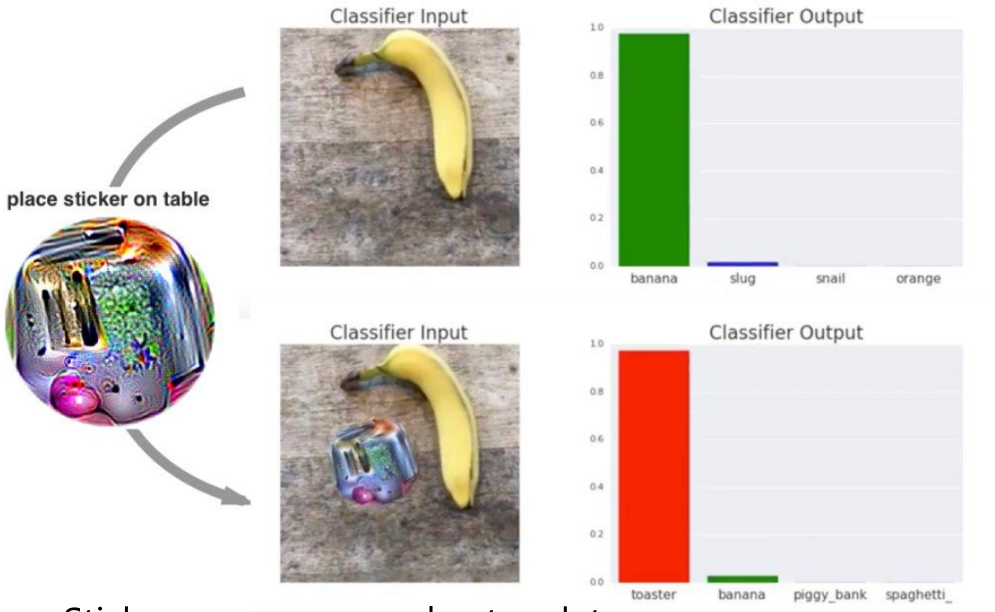


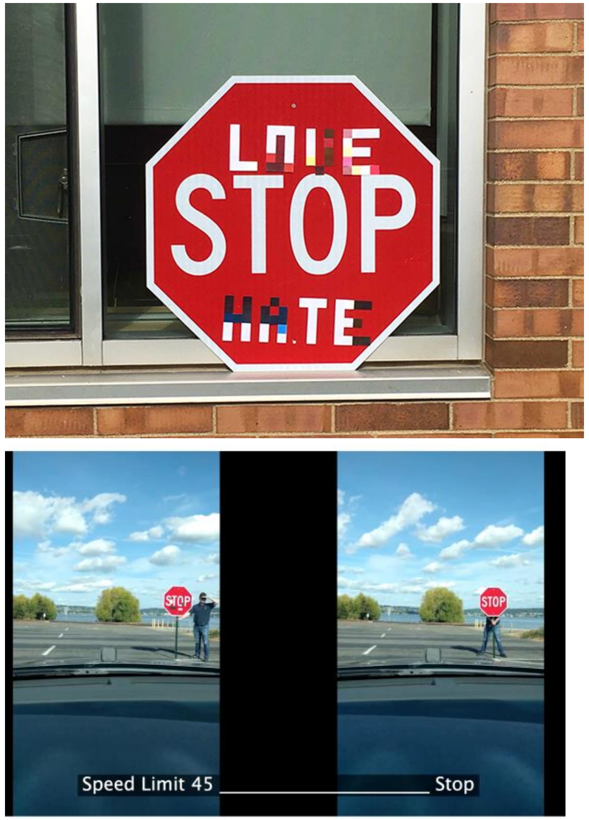
Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

Nguyen, Yosinksi et al. 2015

Crimes *on* AI | also in physical space, ‘adversarial input’ can exploit intrinsic vulnerabilities in AI…

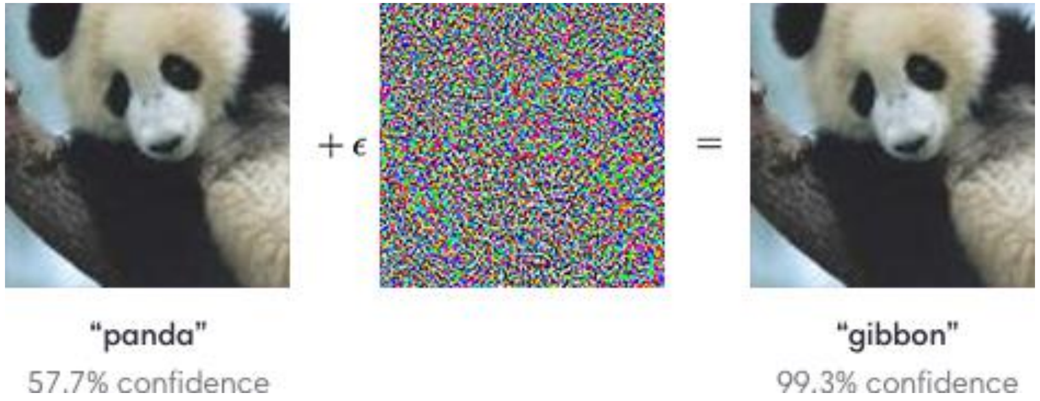


Stickers cause a neural network to misclassify images as a toaster (Brown, Mané, Roy, Abadi, Gilmer, 2017)

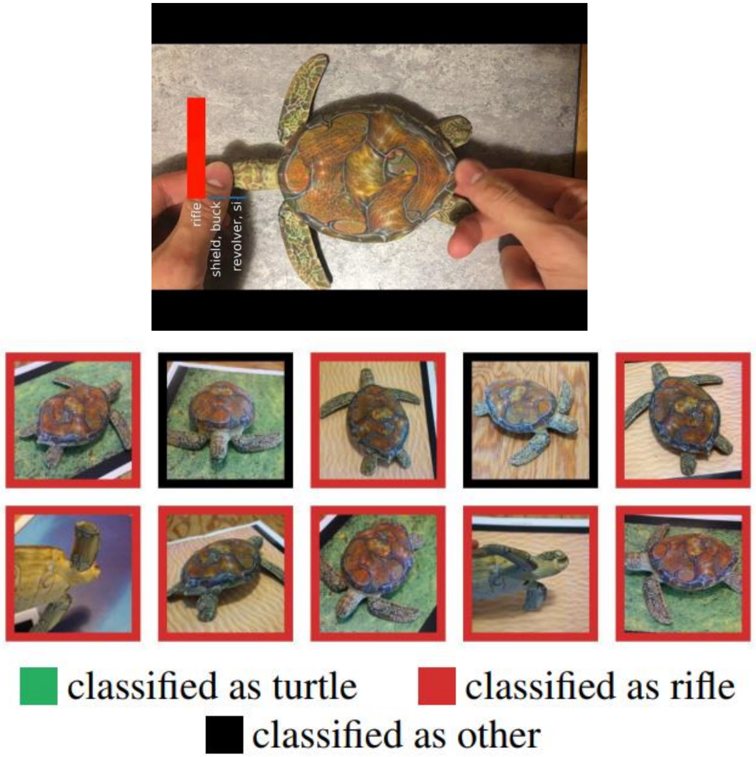


Stickers cause a neural network to misclassify stop signs as ‘speed limit 45’ signs (Evtimov, et al. 2017)

···and can be undetectable to the human eye

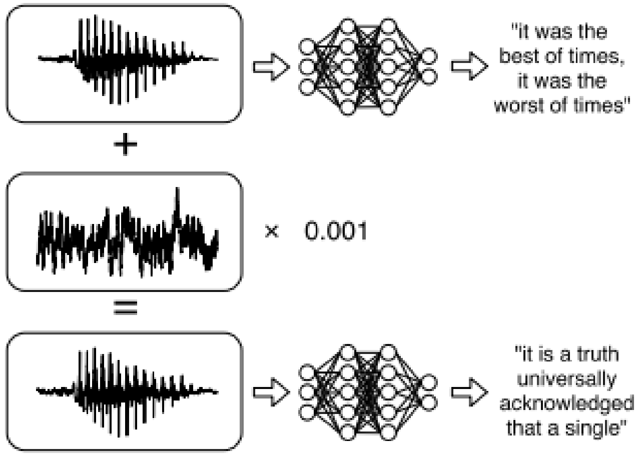


(to humans) imperceptible perturbation of image leads to algorithmic image reclassification (OpenAI, 2017)



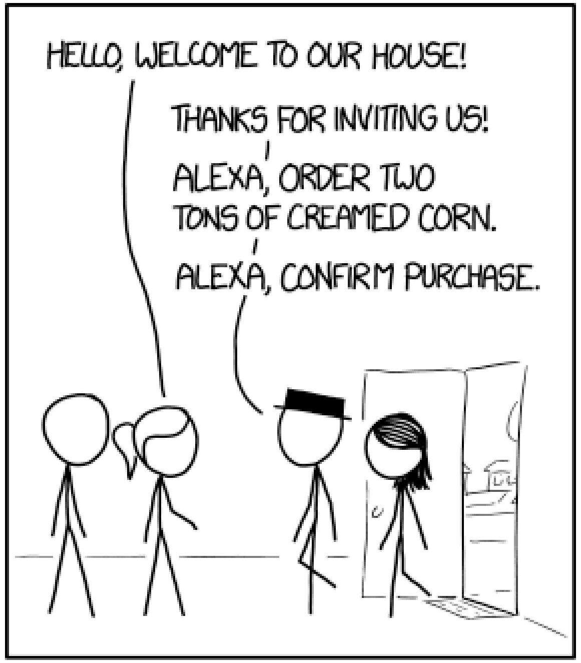
3D-printed turtle, adversarially perturbed [altered] to classify as a 'rifle' at every angle (Athalye et al. 2018)

Crimes *on* AI | (hidden) adversarial input in audio: “Alexa, please transfer money to …”



"it was the best of times, it was the worst of times"
+
× 0.001
=
"it is a truth universally acknowledged that a single"

Figure 1. Illustration of our attack: given any waveform, adding a small perturbation makes the result transcribe as any desired target phrase.



HELLO, WELCOME TO OUR HOUSE!
THANKS FOR INVITING US!
ALEXA, ORDER TWO TONS OF CREAMED CORN.
ALEXA, CONFIRM PURCHASE.

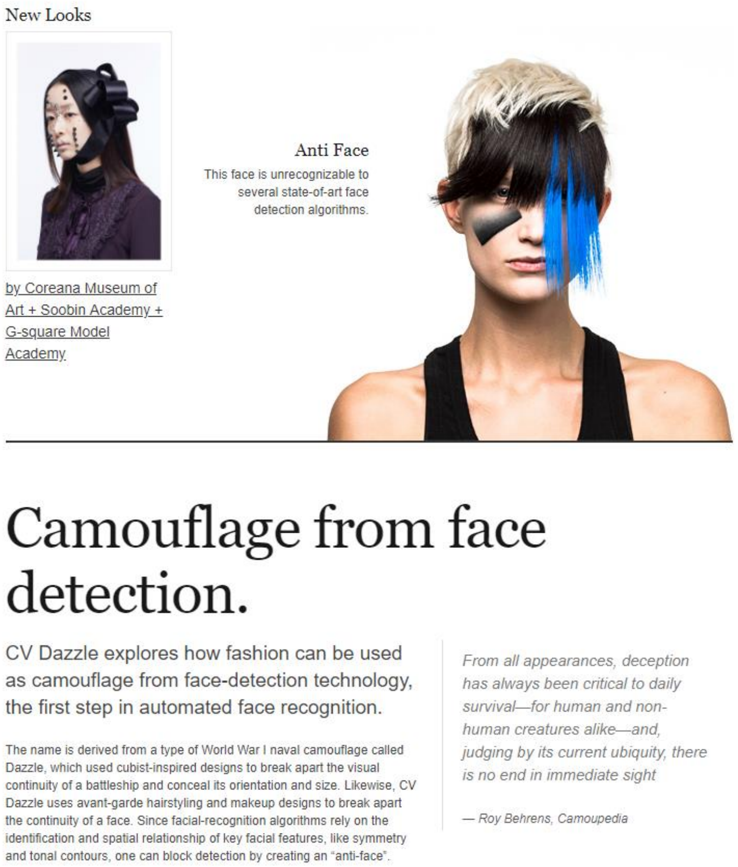
WHEN VISITING A NEW HOUSE, IT'S GOOD TO CHECK WHETHER THEY HAVE AN ALWAYS-ON DEVICE TRANSMITTING YOUR CONVERSATIONS SOMEWHERE.

- “Over the last two years, researchers [⋯] have begun demonstrating that **they can send hidden commands that are undetectable to the human ear to Apple’s Siri, Amazon’s Alexa and Google’s Assistant.** Inside university labs, the researchers have been able to secretly activate the artificial intelligence systems on smartphones and smart speakers, making them dial phone numbers or open websites. In the wrong hands, **the technology could be used to unlock doors, wire money or buy stuff online — simply with music playing over the radio**.”

-[Smith, NYT, 2018]

For examples, see: https://nicholas.carlini.com/code/audio_adversarial_examples/

Resistance against AI surveillance? | 'dazzle fashion' against facial recognition



New Looks

by Coreana Museum of Art + Soobin Academy + G-square Model Academy

Anti Face

This face is unrecognizable to several state-of-art face detection algorithms.

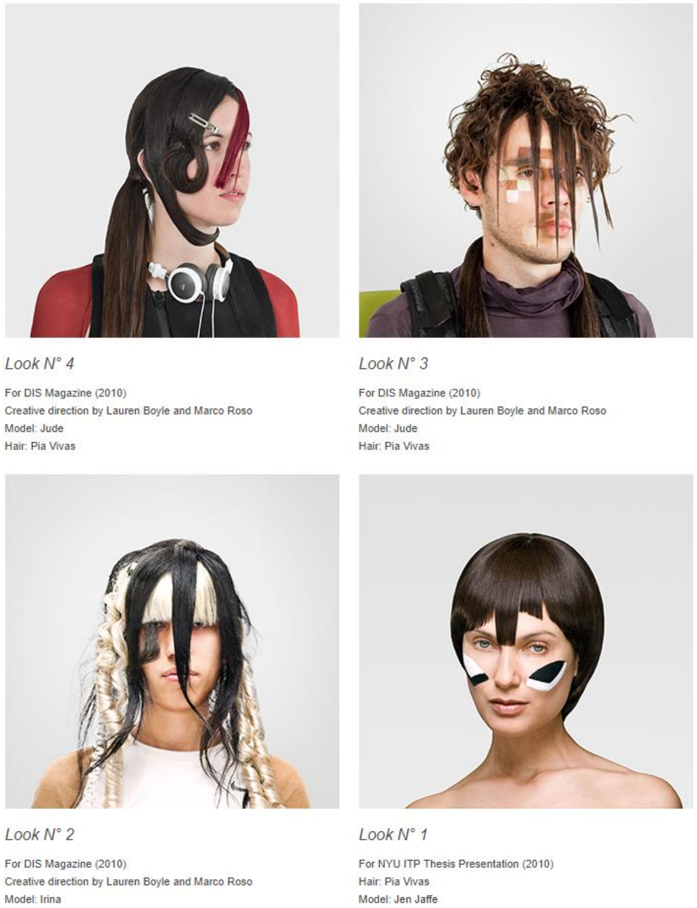
Camouflage from face detection.

CV Dazzle explores how fashion can be used as camouflage from face-detection technology, the first step in automated face recognition.

The name is derived from a type of World War I naval camouflage called Dazzle, which used cubist-inspired designs to break apart the visual continuity of a battleship and conceal its orientation and size. Likewise, CV Dazzle uses avant-garde hairstyling and makeup designs to break apart the continuity of a face. Since facial-recognition algorithms rely on the identification and spatial relationship of key facial features, like symmetry and tonal contours, one can block detection by creating an "anti-face".

> *From all appearances, deception has always been critical to daily survival—for human and non-human creatures alike—and, judging by its current ubiquity, there is no end in immediate sight*
>
> *— Roy Behrens, Camoupedia*

https://cvdazzle.com/



Look N° 4

For DIS Magazine (2010)
Creative direction by Lauren Boyle and Marco Roso
Model: Jude
Hair: Pia Vivas

Look N° 3

For DIS Magazine (2010)
Creative direction by Lauren Boyle and Marco Roso
Model: Jude
Hair: Pia Vivas

Look N° 2

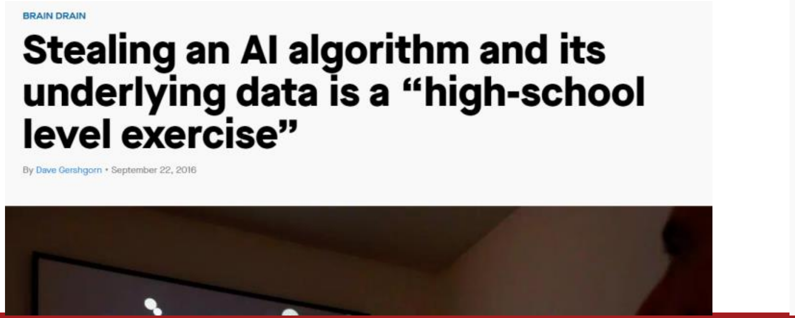
For DIS Magazine (2010)
Creative direction by Lauren Boyle and Marco Roso
Model: Irina

Look N° 1

For NYU ITP Thesis Presentation (2010)
Hair: Pia Vivas
Model: Jen Jaffe

Crimes *on* AI | ‘How to Steal an AI’

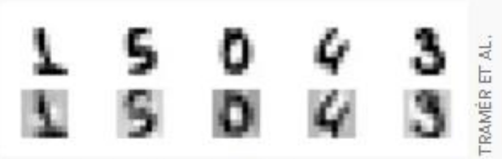
- Copying a publicly embedded AI [e.g. Amazon book recommendation algorithm] simply by ‘asking it a lot of questions’ & training your own neural network on its input/output..
- The complexity of the algorithm determines how **hard it is to steal**. Simple yes-or-no algorithms can be copied in just 41 queries, less than $0.10 under Google’s payment structure. Complex neural networks **on average took 108,200 queries**, but achieved more than 98% accuracy against the original algorithm. –[Tramer, Zhang et al. 2016]
- Ability to steal AI products from companies like Microsoft and IBM; but also small, single-machine-learning API startups. Also helps in developing adversarial input.



BRAIN DRAIN

Stealing an AI algorithm and its underlying data is a “high-school level exercise”

By Dave Gershgorn • September 22, 2016

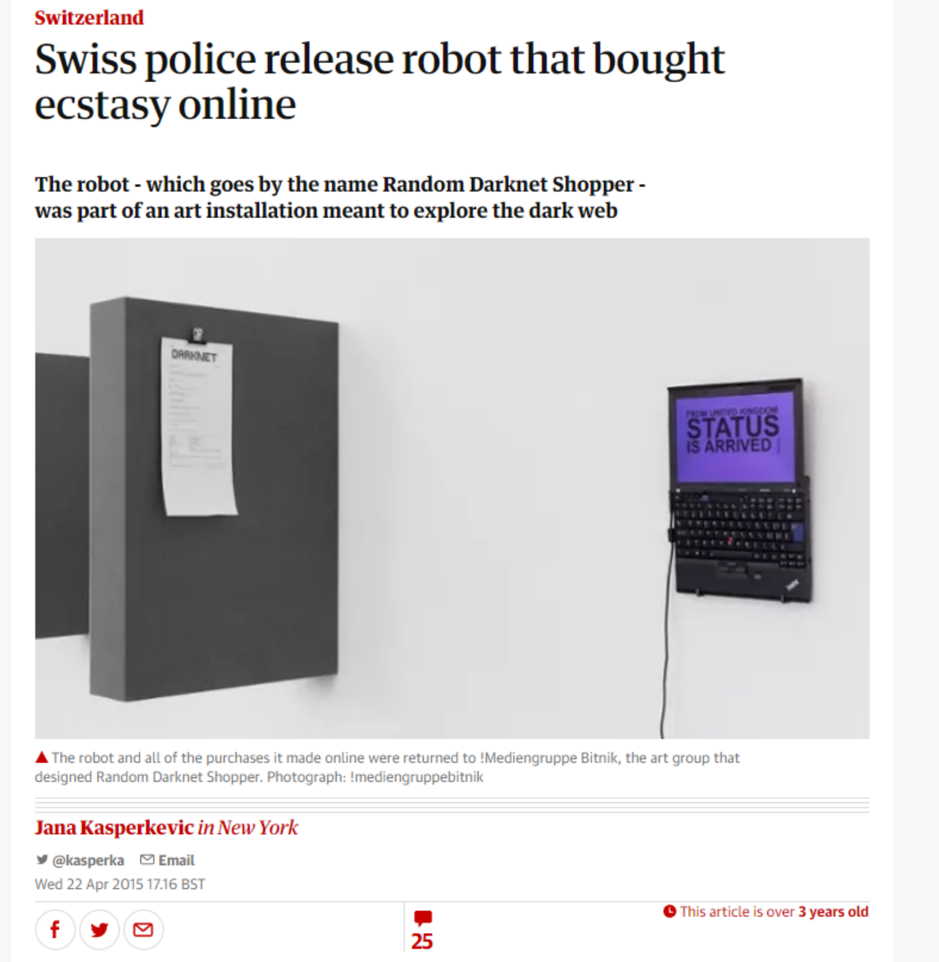


TRAMÈR ET AL.

The original training data (top) against the recovered data from the stolen algorithm (bottom).

- What is AI?
- **Criminal uses of AI**
  - Crimes *with* AI: scalable cyberattacks; DeepFakes
  - Crimes *on* AI: a new attack surface; adversarial input; 'AI theft'
  - **Crimes *by* AI: 'algorithmic entities' as criminal intermediary?**
- Policing use of AI
- Issues and opportunities
- Conclusions, overall takeaways

Crimes *by* AI? | AI as white-collar criminal ‘shield’?



Switzerland

Swiss police release robot that bought ecstasy online

The robot - which goes by the name Random Darknet Shopper - was part of an art installation meant to explore the dark web

▲ The robot and all of the purchases it made online were returned to !Mediengruppe Bitnik, the art group that designed Random Darknet Shopper. Photograph: !mediengruppebitnik

Jana Kasperkevic *in New York*

@kasperka Email

Wed 22 Apr 2015 17.16 BST

This article is over **3 years old**

25

Taking that further: legally recognized ‘Algorithmic Entities’?

- It may be possible today to establish a limited liability company under current US law, and put an algorithmic entity solely in charge of it [Bayern 2015], functionally bestowing legal personhood on an AI
- *“Initiators can **limit their civil and criminal liability** for acts of their algorithms by transferring the algorithms to **entities** and surrendering control at the time of the launch.75 For example, the initiator might specify a general goal, such as maximizing financial return, and leave it to the algorithm to decide how to do that. **If the algorithm later directed the commission of a crime, prosecutors may be unable to prove the intent necessary to convict the initiator of that crime (as opposed to the lesser charge of reckless initiation).**”* [-LoPucki 2018]

Summary: AI and Crime

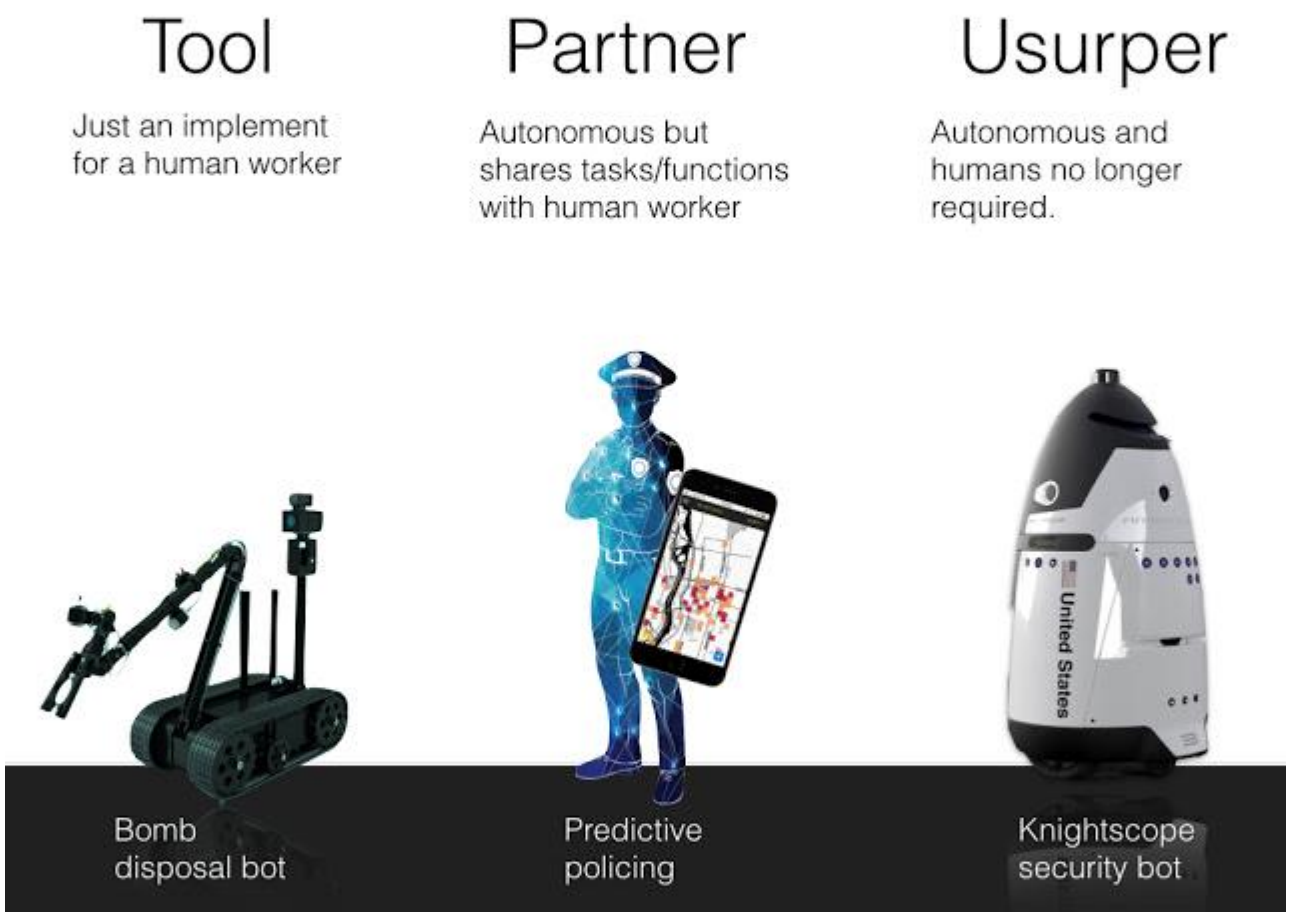
- **Crimes *with* AI**: ability to scale up existing crimes (e.g. cyberattacks); introduces new attacks (e.g. generate fake material)
- **Crimes *on* AI**: AI as a new attack surface, vulnerable to theft or to being spoofed or even controlled through (audio/image) adversarial input.
- **Crimes *by* AI:** 'Algorithmic entities' operating as intermediary for others render identification and proof of criminal involvement or intent difficult

- The danger of 'Artificial Stupidity' as new vulnerability, widely dispersed in society

- What is AI?
- Criminal uses of AI
- **Policing use of AI**
  - Monitoring: scaling up
  - Prediction: plagued by bias
  - Sentencing: better accuracy?
  - Structure: Automating situational crime prevention?
- Issues and opportunities
- Conclusions, overall takeaways

- What is AI?
- Criminal uses of AI
- **Policing use of AI**
  - **Monitoring: scaling up**
  - Prediction: plagued by bias
  - Sentencing: better accuracy?
  - Structure: Automating situational crime prevention?
- Issues and opportunities
- Conclusions, overall takeaways

AI in policing: tool, partner, usurper?



Tool

Just an implement
for a human worker

Partner

Autonomous but
shares tasks/functions
with human worker

Usurper

Autonomous and
humans no longer
required.

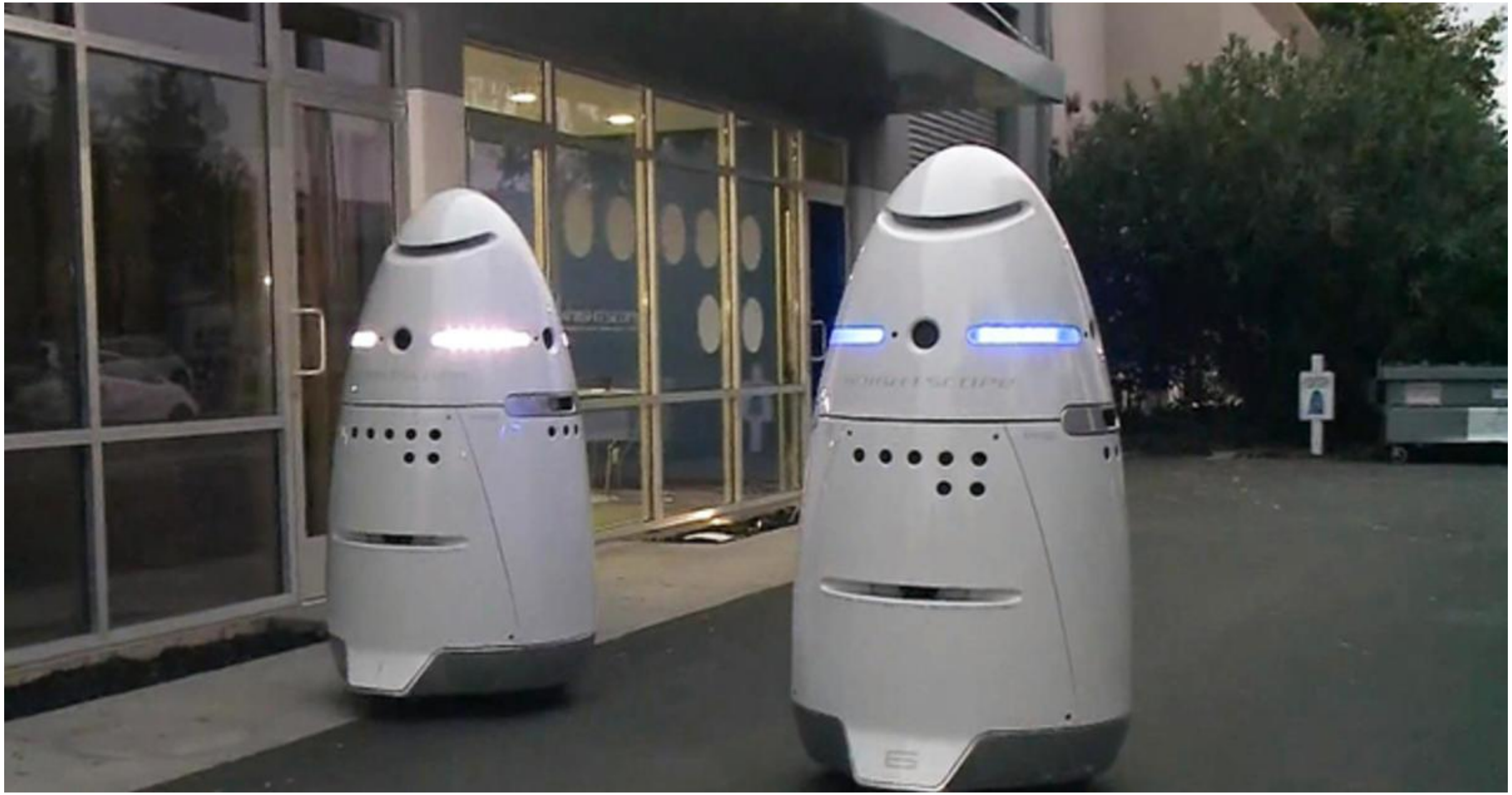
Bomb
disposal bot

Predictive
policing

Knightscope
security bot

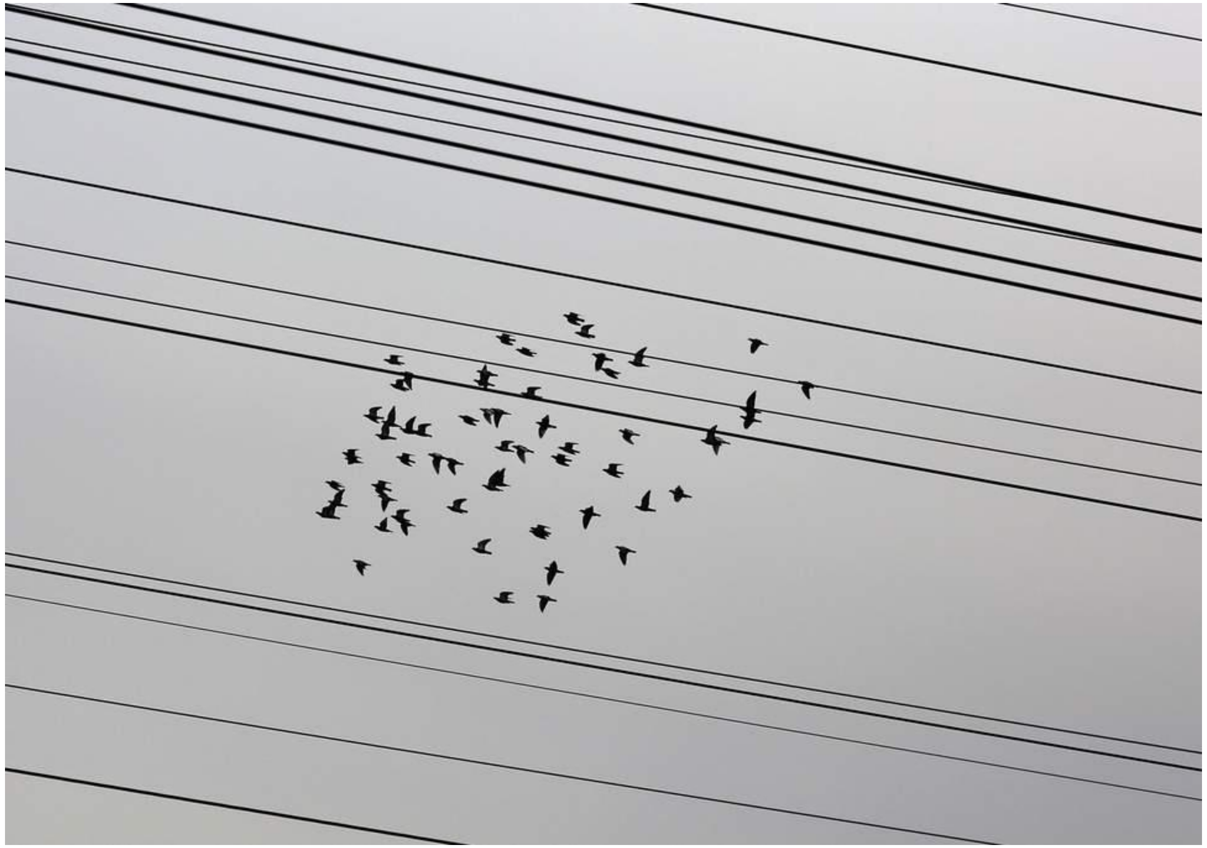
[Danaher 2018]

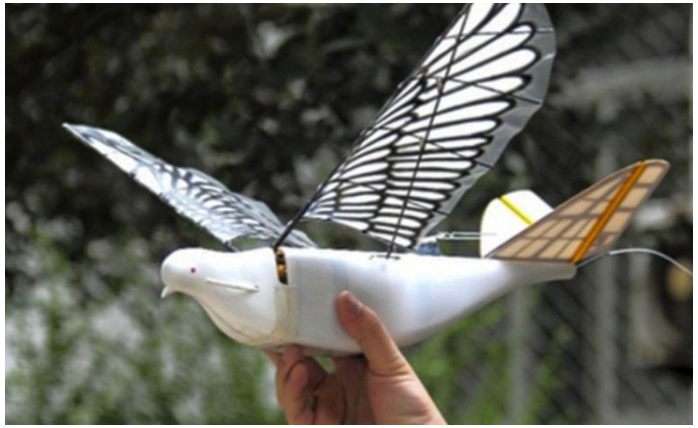
Monitoring AI: real ‘robocops’? [Also cf. private sector policing].



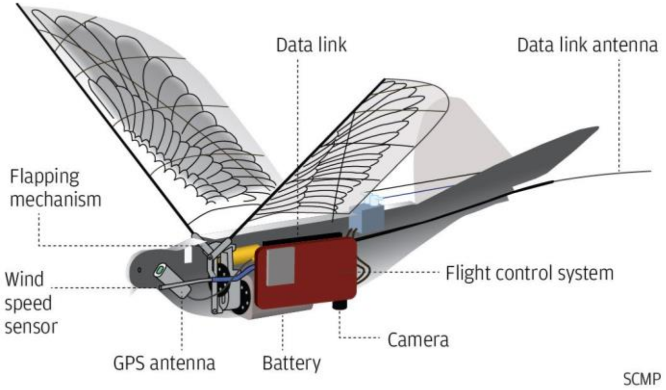
Monitoring: also surveillance in plain sight…

- China’s bird drones (code-named ‘Dove’) operational in 5 provinces





Eye in the sky

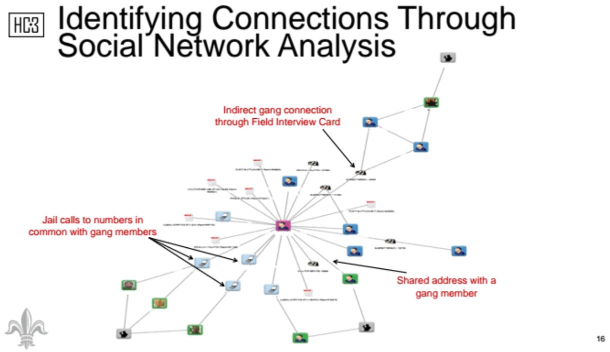


Cf. comprehensive ‘social’ surveillance—Chinese Social Credit system

- By 2020: full implementation of social credit score system
- **Input**: full social profile, including location, friends, health records, insurance, private messages, financial position, gaming duration, smart home statistics, preferred newspapers, shopping history, and dating behaviour
- **Results**: flight ban, exclusion from private schools, slow internet connection, exclusion from high prestige work, exclusion from hotels, and registration on a public blacklist.

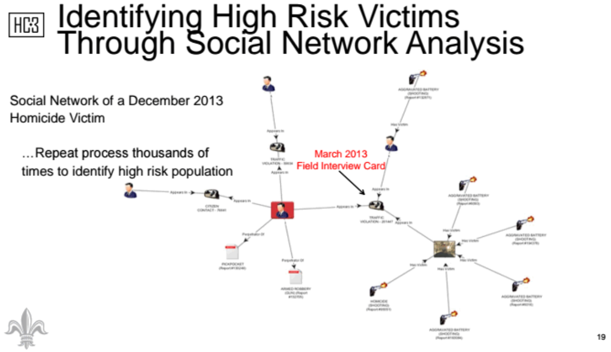
- What is AI?
- Criminal uses of AI
- **Policing use of AI**
  - Monitoring: scaling up
  - **Prediction: plagued by bias**
  - Sentencing: better accuracy?
  - Structure: Automating situational crime prevention?
- Issues and opportunities
- Conclusions, overall takeaways

Predictive Policing



HC3 Identifying Connections Through Social Network Analysis

Indirect gang connection through Field Interview Card
Jail calls to numbers in common with gang members
Shared address with a gang member
16

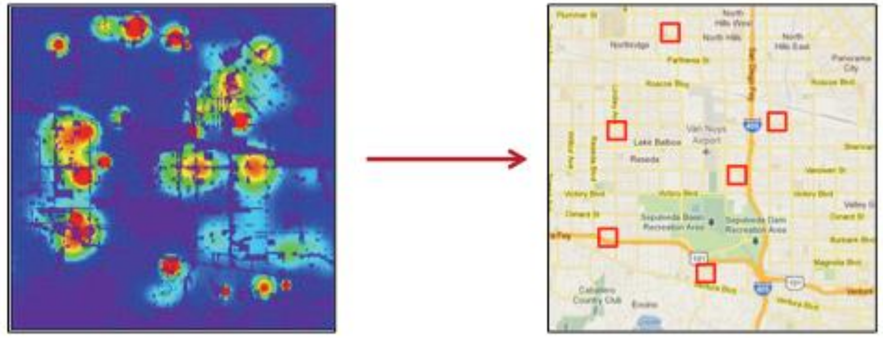


HC3 Identifying High Risk Victims Through Social Network Analysis

Social Network of a December 2013 Homicide Victim

…Repeat process thousands of times to identify high risk population

- (secretly) used by *Palantir* in NOLA (Operation LASER) until March 2018
- Also PredPol software, which inputs three variables of crime (where, when, what type) to calculate ‘hotspots’
- Does it work? RAND report suggests no statistically significant crime reduction…
- ⋯but system enforces racially biased policing patterns and practices



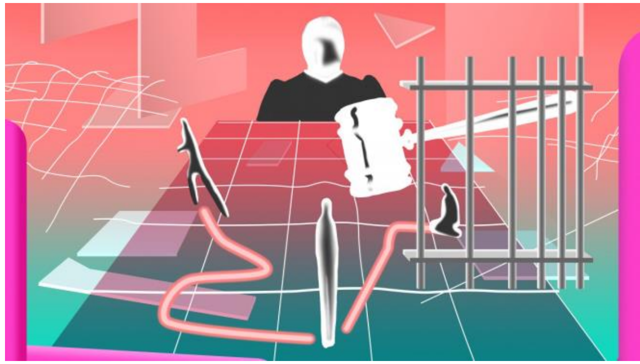
TACTICAL AMBIGUITY
rear-view mirror heat map

TACTICAL CLARITY
forward-looking PredPol boxes

- What is AI?
- Criminal uses of AI
- **Policing use of AI**
  - Monitoring: scaling up
  - Prediction: plagued by bias
  - **Sentencing: better accuracy?**
  - Structure: Automating situational crime prevention?
- Issues and opportunities
- Conclusions, overall takeaways

Algorithmic sentencing: promise and peril

- Use of algorithms to determine e.g. sentencing, bail decisions
- Goal to reduce bias and increase accuracy: (some) algorithms better at predicting risk or recidivism than human judges:
  - In deciding on whether a defendant should await trial in jail, a NBER algorithm could cut **crime by defendants awaiting trial by as much as 25 percent without changing the numbers of people waiting in jail**.
  - Or it could **reduce the jail population awaiting trial by more than 40 percent** while leaving the **crime rate by defendants unchanged**.
- But: risks from bias? (COMPAS)
- Shift from concept of law to ‘regulatory instrumentalism’?



- What is AI?
- Criminal uses of AI
- **Policing use of AI**
  - Monitoring: scaling up
  - Prediction: plagued by bias
  - Sentencing: better accuracy?
  - **Structure: automating situational crime prevention?**
- Issues and opportunities
- Conclusions, overall takeaways

Technology, and from normative to non-normative 'regulatory modalities'… [Lessig, 2001, Browsword 2016]

Golf Club: how to solve golf cart flower-surfing?

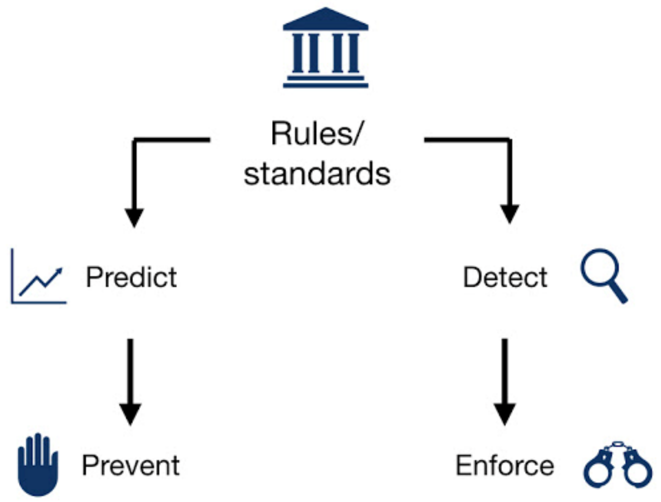
Normative:

- **Norms** (*proper members don't do that; you'll get shunned*)
- **Law** *(…we now have a rule that says you'll get fined if we catch you doing it···)*
- **Law+(Technology-enforced law)** *(…and we have cameras to ensure we catch you doing it.)*

Non-normative:

- Technological management (*built-in GPS shuts off your cart when approaching the flowers)*

‘Technological management’… [Browsword 2016]



[Danaher 2018]

Automation drives shift…

from ‘**Detect and Enforce**’ to ‘**Predict & Prevent**’

Cf. ‘Situational Crime Prevention’ (SCP)

…the ‘Disneyfication’ of policing in future smart cities?

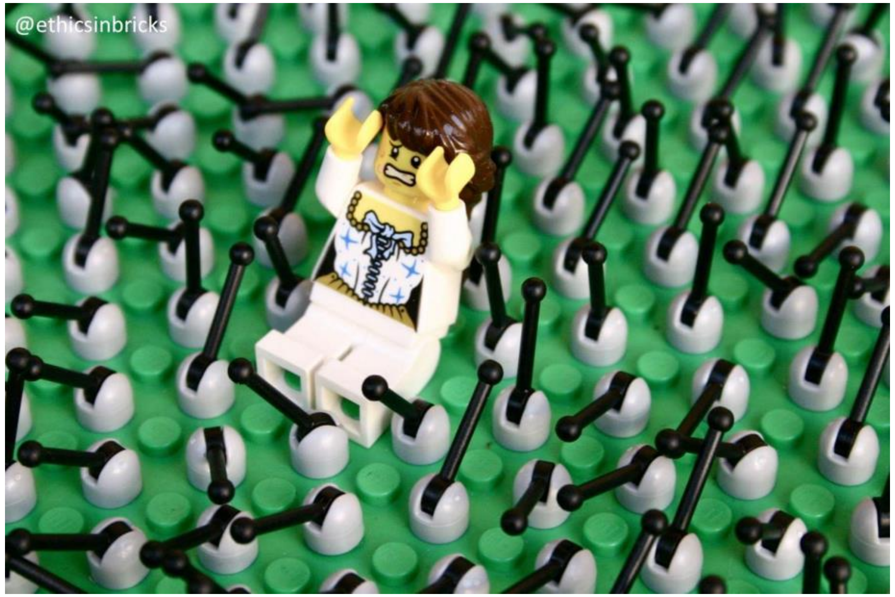
”[Disneyland] anticipates and prevents possibilities for disorder through constant instructions to visitors, physical barriers that both guide and limit visitors’ movements, and through “omnipresent” employees who detect and correct the smallest errors (Shearing & Stenning 1985: 301). ***None of the costumed characters nor the many signs, barriers, lanes, and gardens feel coercive to visitors. Yet through constant monitoring, prevention, and correction embedded policing is part of the experience…”***

[Joh 2018]

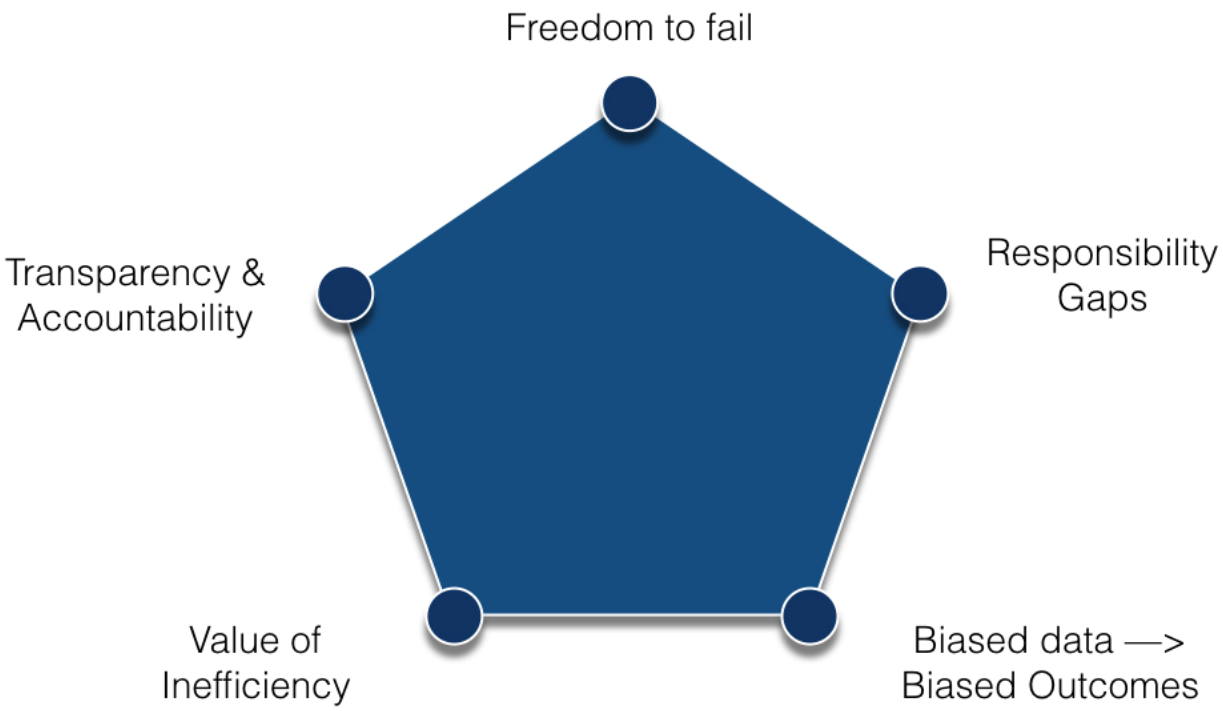
Summary: AI and Policing

- Set to drive shift in policing…
  - **Monitoring**: …from visible to ‘invisible’ (yet known to be all-pervasive)
  - **Prediction**: …from ‘detect and enforce’, to ‘predict and prevent’
  - **Sentencing**: …from rule of law to ‘regulatory instrumentalism’
  - **Structure**: …from ‘normative’, to ‘non-normative’ technological management / automated Situational Crime Prevention approach

- What is AI?
- Criminal uses of AI
- Policing use of AI
- **Issues and opportunities**
  - **Overall considerations**
  - Does AI surveillance need to be intrusive?
- Conclusions, overall takeaways



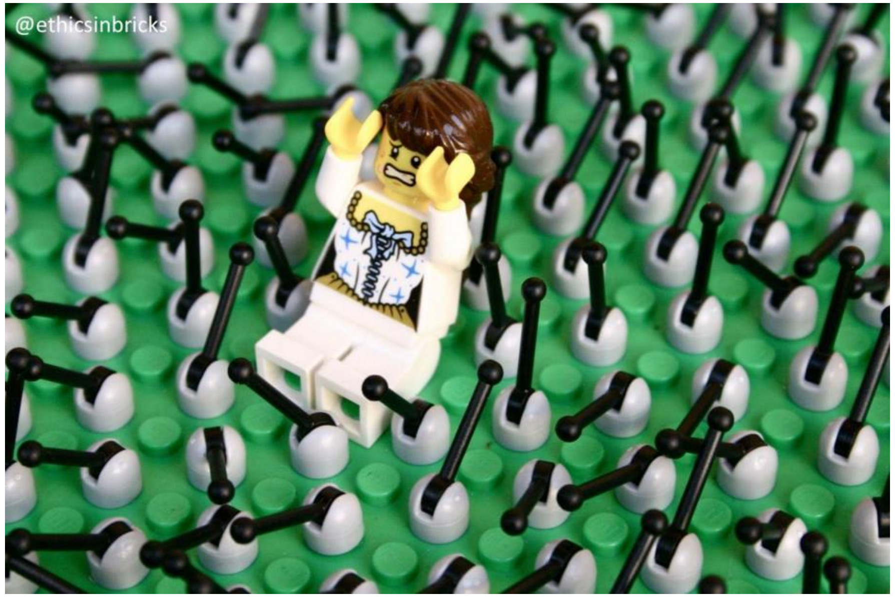
Five Considerations (cf. Danaher 2018)



[Danaher 2018]

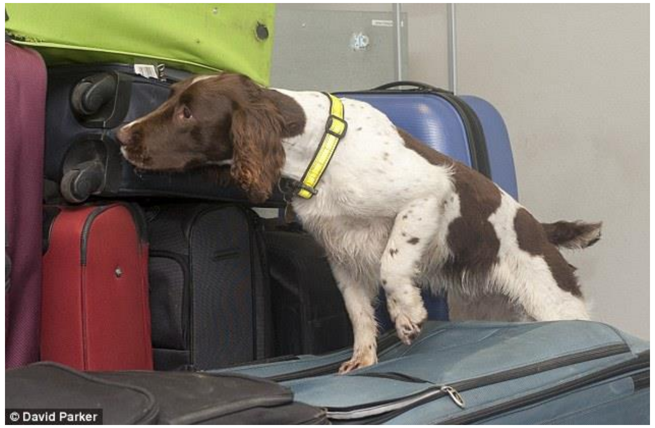
Garbage-in-garbage-out!

- What is AI?
- Criminal uses of AI
- Policing use of AI
- **Issues and opportunities**
  - Overall considerations
  - **Does AI surveillance need to be intrusive?**
- Conclusions, overall takeaways



Does surveillance need to be intrusive? - Possible advantages of automation [Garfinkel 2018]

- Discussions of surveillance often assume core trade-offs:
  - Privacy vs. security (*'to find the minority of bad actors, we must have access to everyone')*
  - Accountability vs. security (*'if we make a search protocol public, bad actors can alter their behaviour'*)
- But are these tradeoffs intractable?
  - Bomb/drug-sniffing dogs reduce the need for human search of all luggage; only 'flag' suspect behavior.
  - *United States V. Place* – ruled that because drug dogs exclusive detect odor of narcotics, they are not considered a 'search' or invasion of privacy.


© David Parker

Does surveillance need to be intrusive? - Possible advantages of automation [Garfinkel 2018]

But automation…

"Can act as a "screen" between data and humans

- *Can judge whether human search/access is warranted; can redact sensitive information (e.g. automatic face blurring)*

Can increase accuracy:

- *Less need for searches/violations of privacy based on false positives*

In certain regards, more predictable and less opaque than humans

- *Not a complete "black box," as humans are; clear audit logs; less likely to engage in certain abuses (LOVEINT, extortion, etc.)*

Makes auditing more "scalable" :

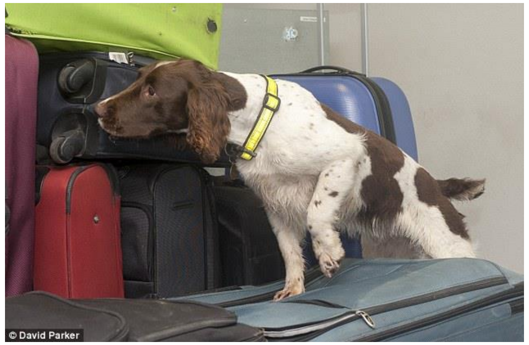
- *Easier to audit one piece of software, than many humans*

Easier to associate with summary statistics (e.g. accuracy rate)*—no 'plausible deniability' of bias. There is a reason why we Pro Publica was able to 'expose' COMPAS"*

Homomorphically Encrypted AI: privacy-preserving, less intrusive surveillance? [Trask 2017]

- Encrypted neural network, which finds, reveals only data that matches pre-determined (but encrypted—ergo hidden) criteria for illegal/suspect activity, and does not reveal any other public data.

- *“Say we’re talking about the surveillance cameras of the future, which come up with encrypted images. Why would we want to do that? [···]* ***If you’re looking for a suspect, you might be interested in doing some computations on an encrypted image, to match [only] to the subject.****”*



© David Parker

- What is AI?
- Criminal uses of AI
- Policing use of AI
- Issues and opportunities
- **Conclusions, overall takeaways**

In sum

Now that we're in the future… What *does* it look like?

I. **AI is not magic**: we can [and must!] engage with the technology and its complexity, if we are to hold it accountable

II. **Criminal uses of AI**: scaling up existing crimes; enabling novel crimes—amplified by prevalence of algorithmic 'artificial stupidity'

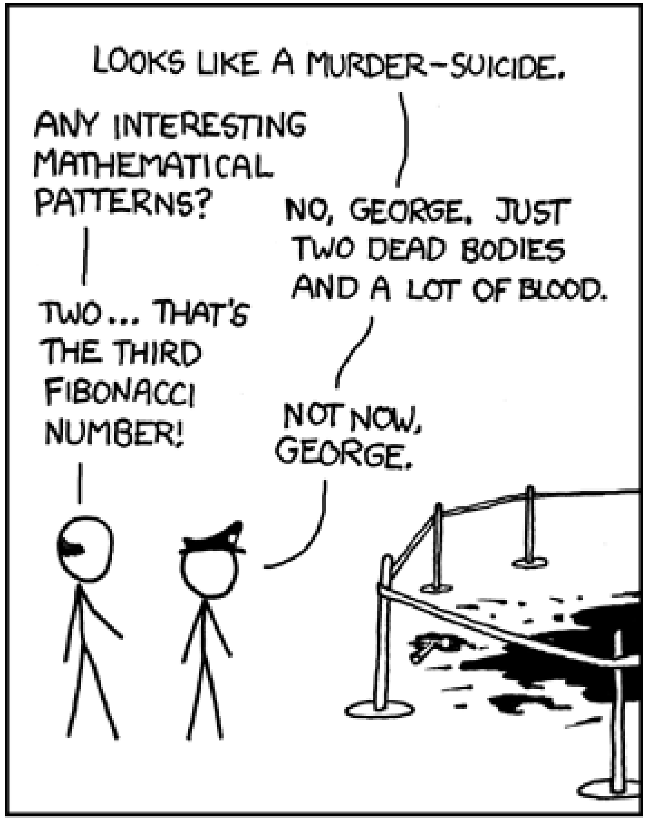
III. **Policing use of AI**: more diffuse monitoring; structural shift towards prediction and prevention

IV. **AI is not magic**: garbage-in-garbage-out; we have to *choose* the right data.

V. **AI is not (black) magic:** …it *could* also help us reduce the intrusiveness of policing and surveillance

Thanks!

- Matthijs.Maas@jur.ku.dk


LOOKS LIKE A MURDER-SUICIDE.
ANY INTERESTING MATHEMATICAL PATTERNS?
NO, GEORGE. JUST TWO DEAD BODIES AND A LOT OF BLOOD.
TWO... THAT'S THE THIRD FIBONACCI NUMBER!
NOT NOW, GEORGE.

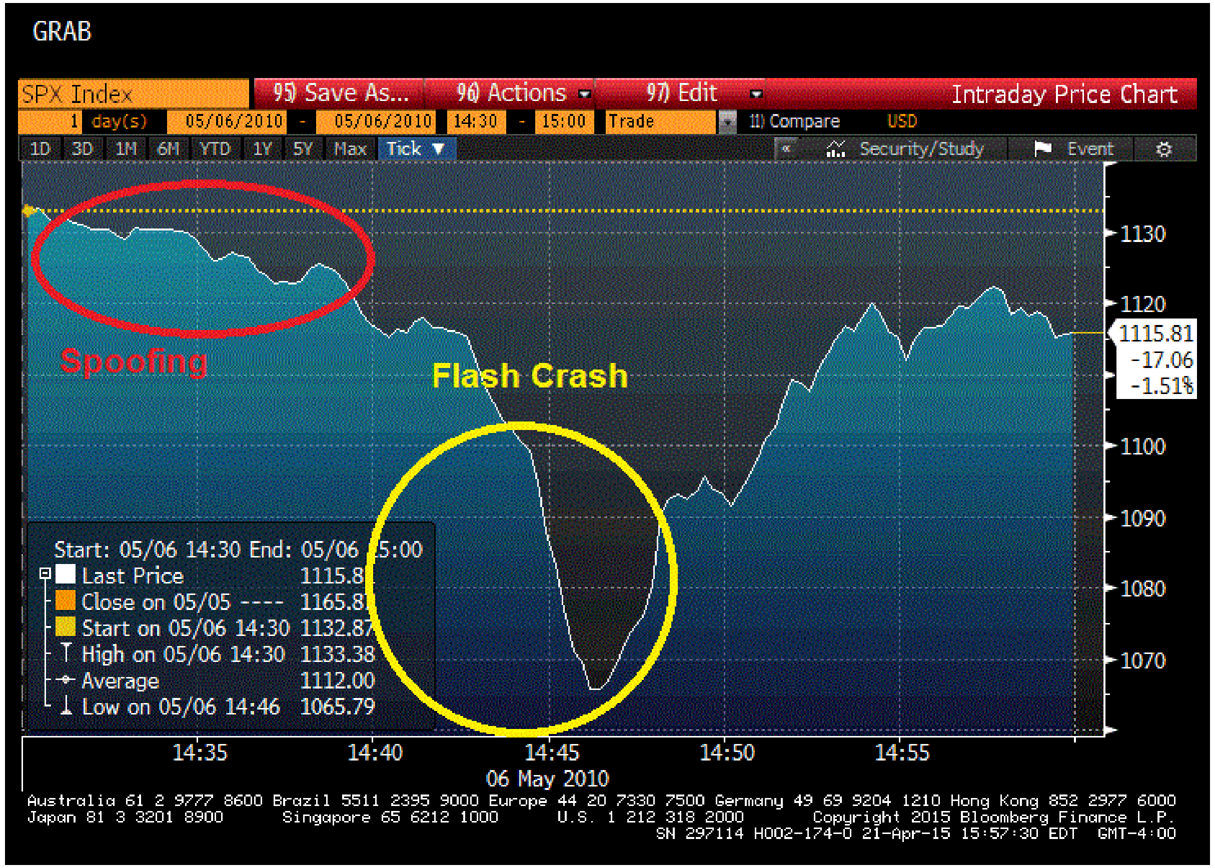
WHEN MATHNET SHUT DOWN, THE OFFICERS HAD TROUBLE REINTEGRATING INTO THE REGULAR L.A.P.D.

Assorted readings:

- Trask, I. (2017) 'Safe Crime Detection: Homomorphic Encryption and Deep Learning for More Effective, Less Intrusive Digital Surveillance' - https://iamtrask.github.io/2017/06/05/homomorphic-surveillance/
- Danaher, John [2018] 'The Automation of Policing: Challenges and Opportunities', *Philosophical Disquisitions*. https://philosophicaldisquisitions.blogspot.com/2018/10/the-automation-of-policing-challenges.html
- Garfinkel, Ben (2018) 'The Future of Surveillance'. https://www.effectivealtruism.org/articles/ea-global-2018-the-future-of-surveillance/

Backup material:

Crimes *on* AI | adversarial input in markets: criminal {?) spoofing can lead to ‘flash crash’



Start: 05/06 14:30 End: 05/06 15:00

⬜ Last Price	1115.8
🟧 Close on 05/05 ----	1165.8
🟨 Start on 05/06 14:30	1132.87
High on 05/06 14:30	1133.38
Average	1112.00
Low on 05/06 14:46	1065.79

Australia 61 2 9777 8600 Brazil 5511 2395 9000 Europe 44 20 7330 7500 Germany 49 69 9204 1210 Hong Kong 852 2977 6000
Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000 Copyright 2015 Bloomberg Finance L.P.
SN 297114 H002-174-0 21-Apr-15 15:57:30 EDT GMT-4:00