OWASP
AppSec Europe
London 2nd-6th June 2018

Seconds out!

When algorithms don’t play nice with our applications and lives.

Etienne Greeff @etienne_greeff

Seconds out!

Etienne Greeff
@etienne_greeff

TOPICS TO COVER

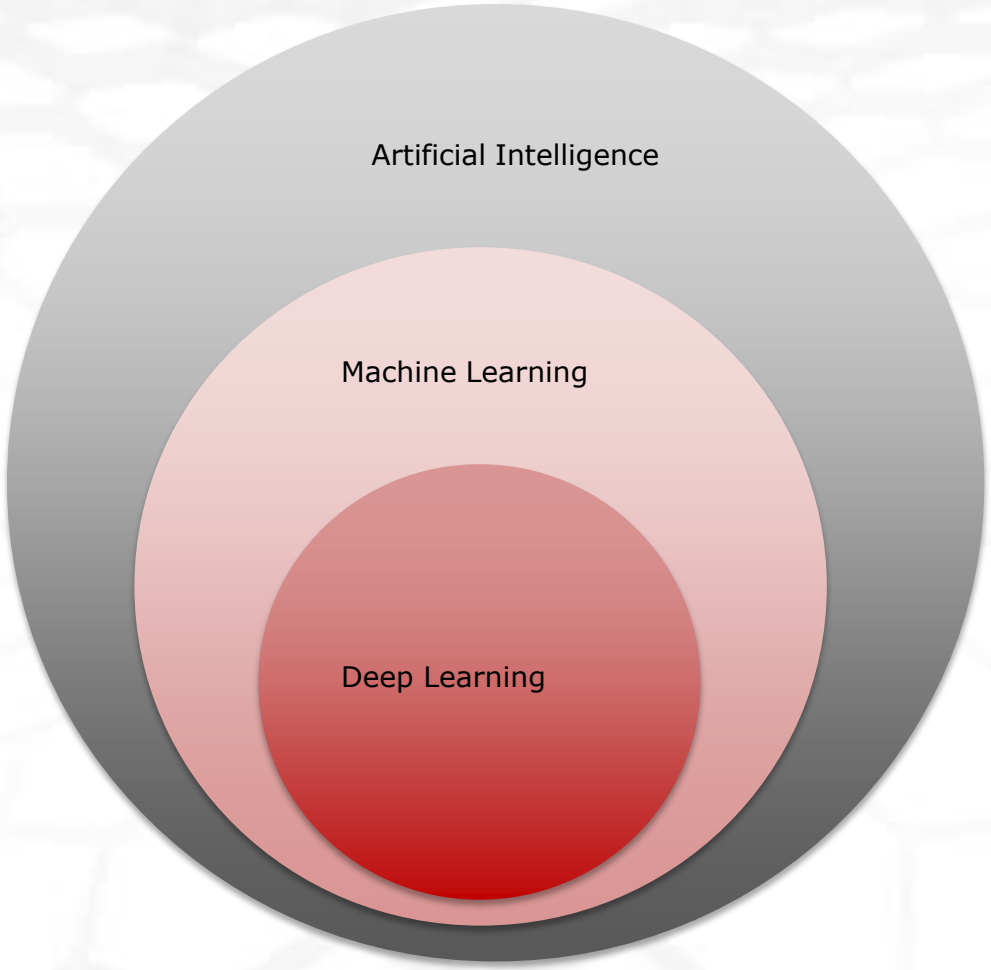
- Introduction
- Important terms
- AI and the security problem
- A tool, not a solution
- The future
- Getting to grips
- Recommendations

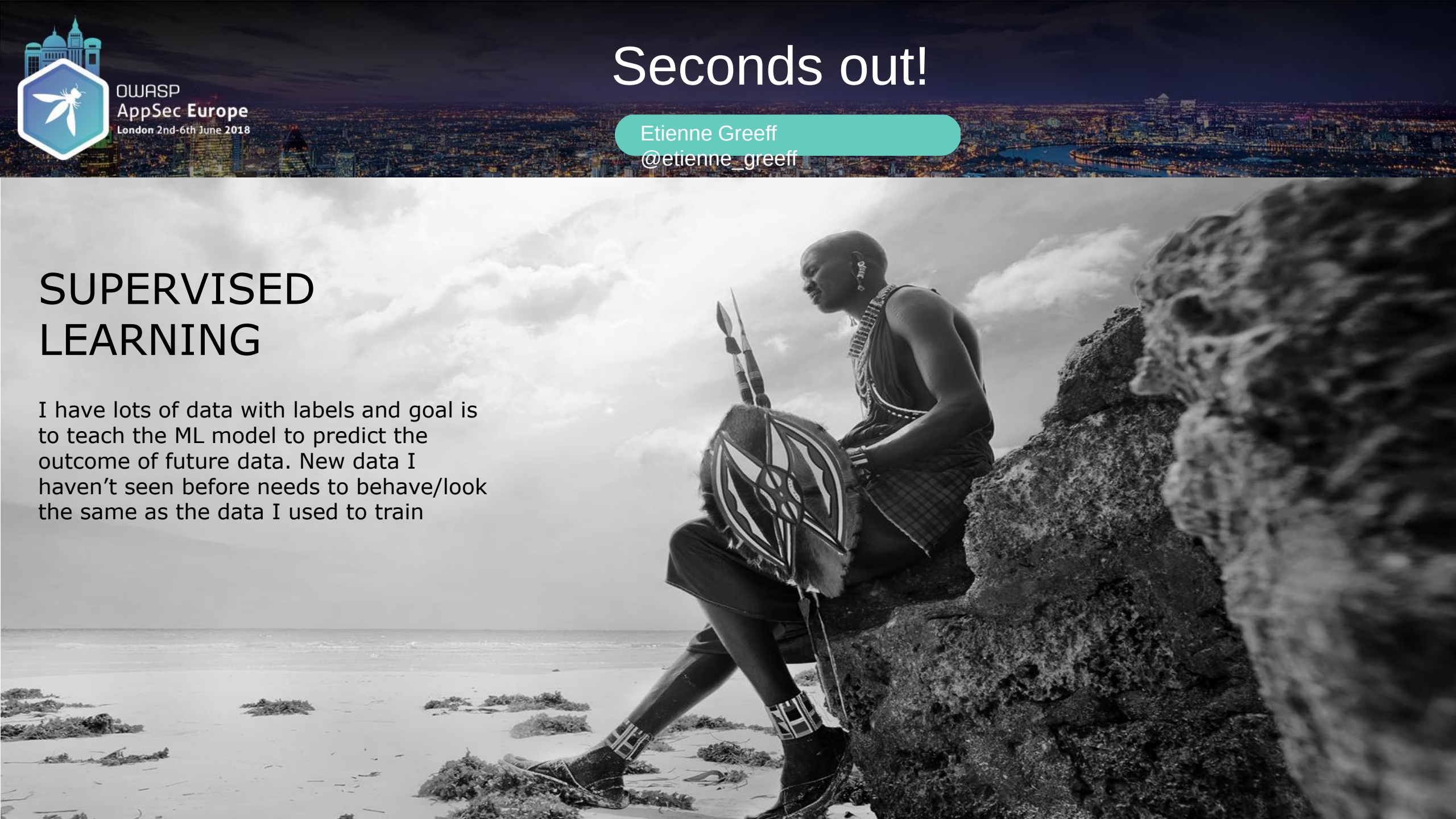
Seconds out!

Etienne Greeff
@etienne_greeff

AI & ML ARE THE SAME YET DIFFERENT?

AI seems to be the encompassing marketing term





Seconds out!

Etienne Greeff
@etienne_greeff

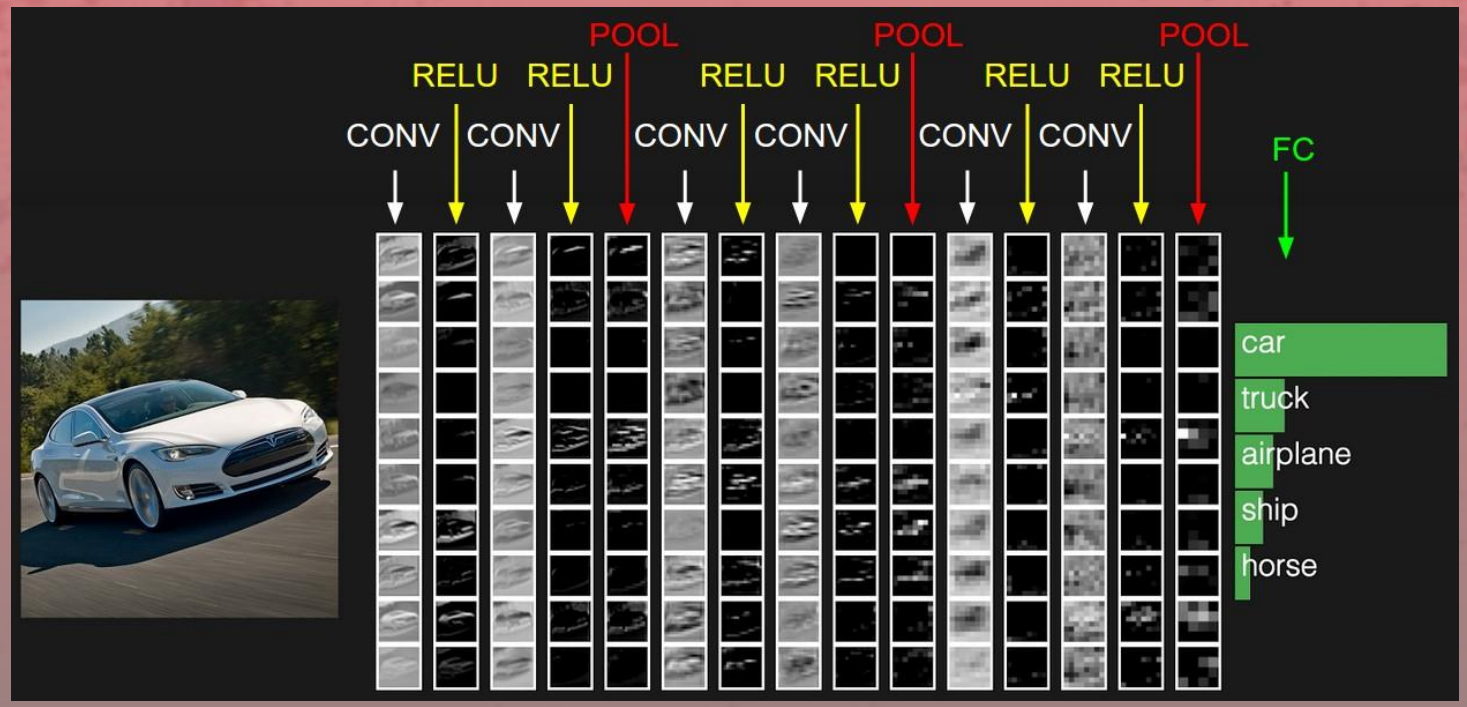
SUPERVISED LEARNING

I have lots of data with labels and goal is to teach the ML model to predict the outcome of future data. New data I haven't seen before needs to behave/look the same as the data I used to train

Seconds out!

Etienne Greeff
@etienne_greeff

EXAMPLE OF SUPERVISED LEARNING: NEURAL NETWORK



Seconds out!

Etienne Greeff
@etienne_greeff

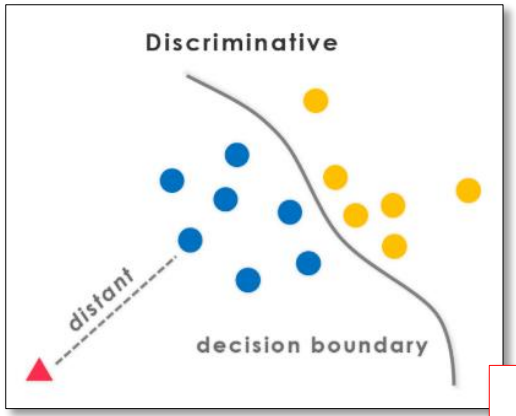
UNSUPERVISED LEARNING

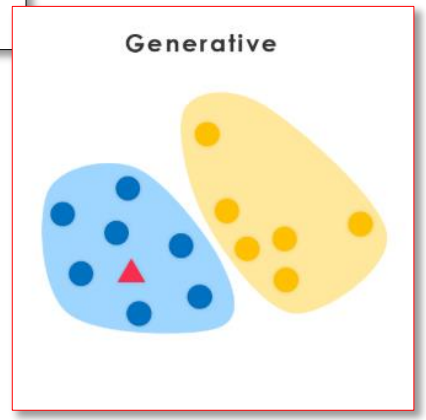
I have lots of data but no labels, so might have lots of logfile entries but don't know which are normal, abnormal, benign or outright malicious

Seconds out!

Etienne Greeff
@etienne_greeff

DISCRIMINATIVE VS GENERATIVE MODELS





Discriminative:

- I am trying to predict the labels by examining the structure in the input data

Generative:

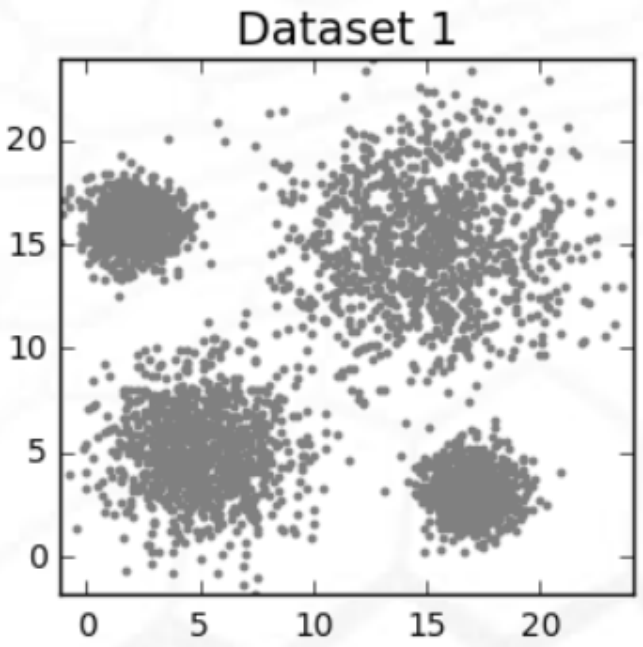
- I am trying to understand the structure in the input data given the labels I am seeing
- Learning involves learning the parameters of the mathematical model but assumes a prior structure
- Can be supervised and unsupervised
- Most of the current generation of defensive tools use generative models

Seconds out!

Etienne Greeff
@etienne_greeff

UNSUPERVISED LEARNING GROUPING SIMILAR POINTS TOGETHER USING THE TWO APPROACHES

- Discriminative – does not rely on prior knowledge of data
- Generative – assumes the data was generated using a known mathematical function (prior)

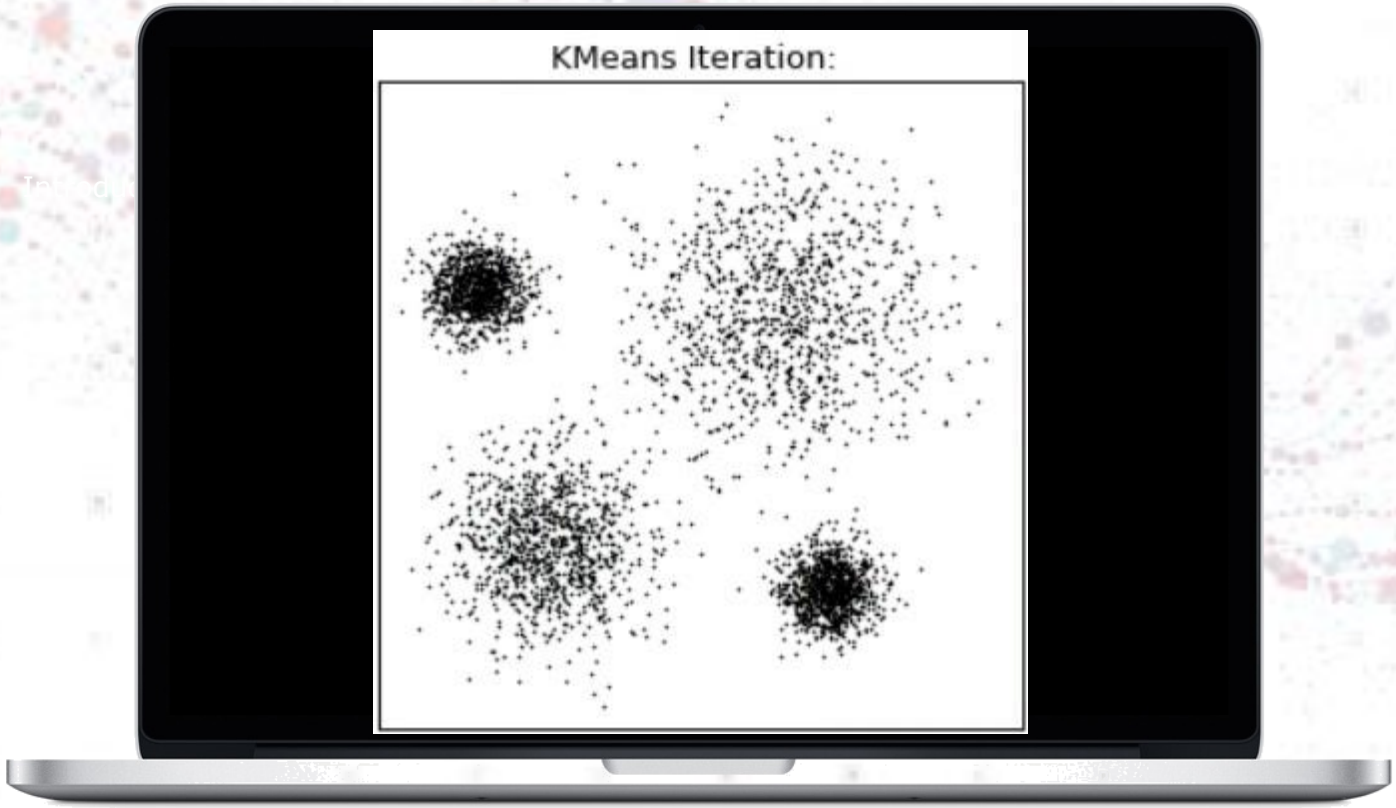


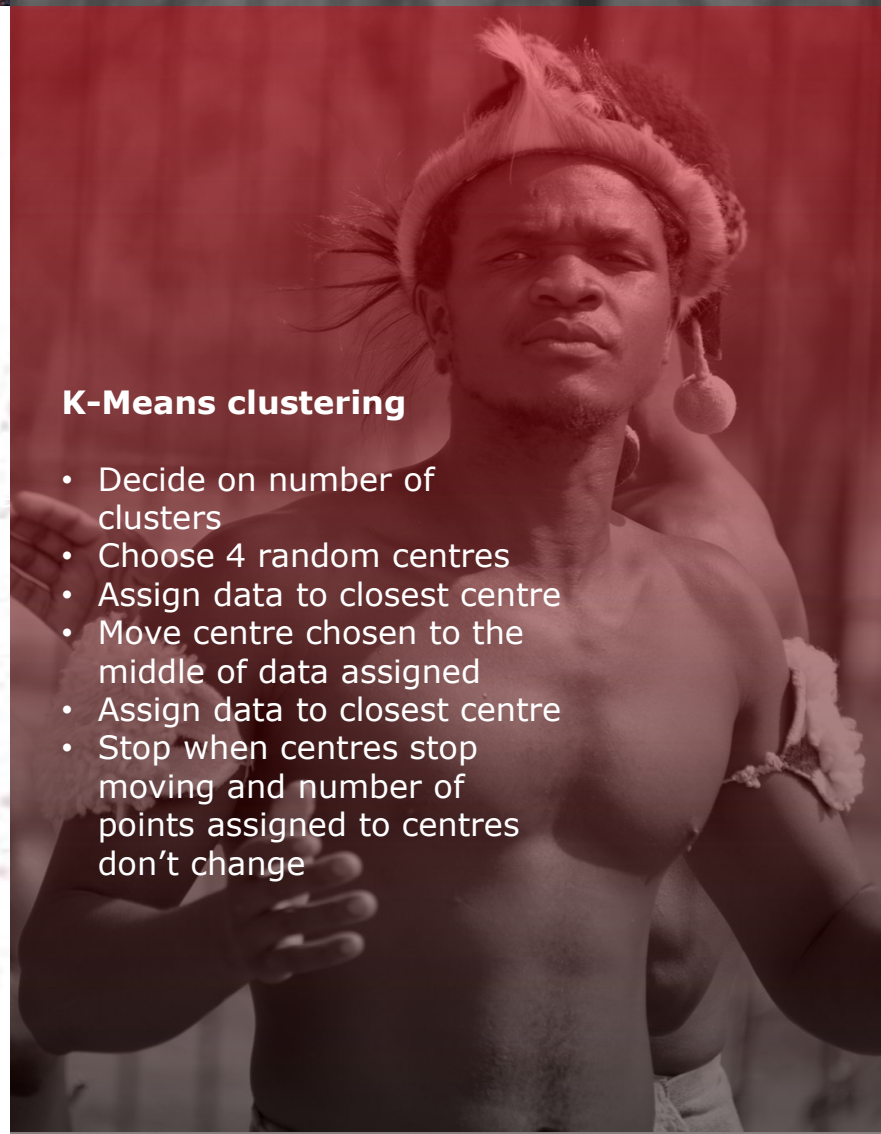
Seconds out!

Etienne Greeff
@etienne_greeff

UNSUPERVISED LEARNING

Using K-means doesn't assume prior knowledge about data





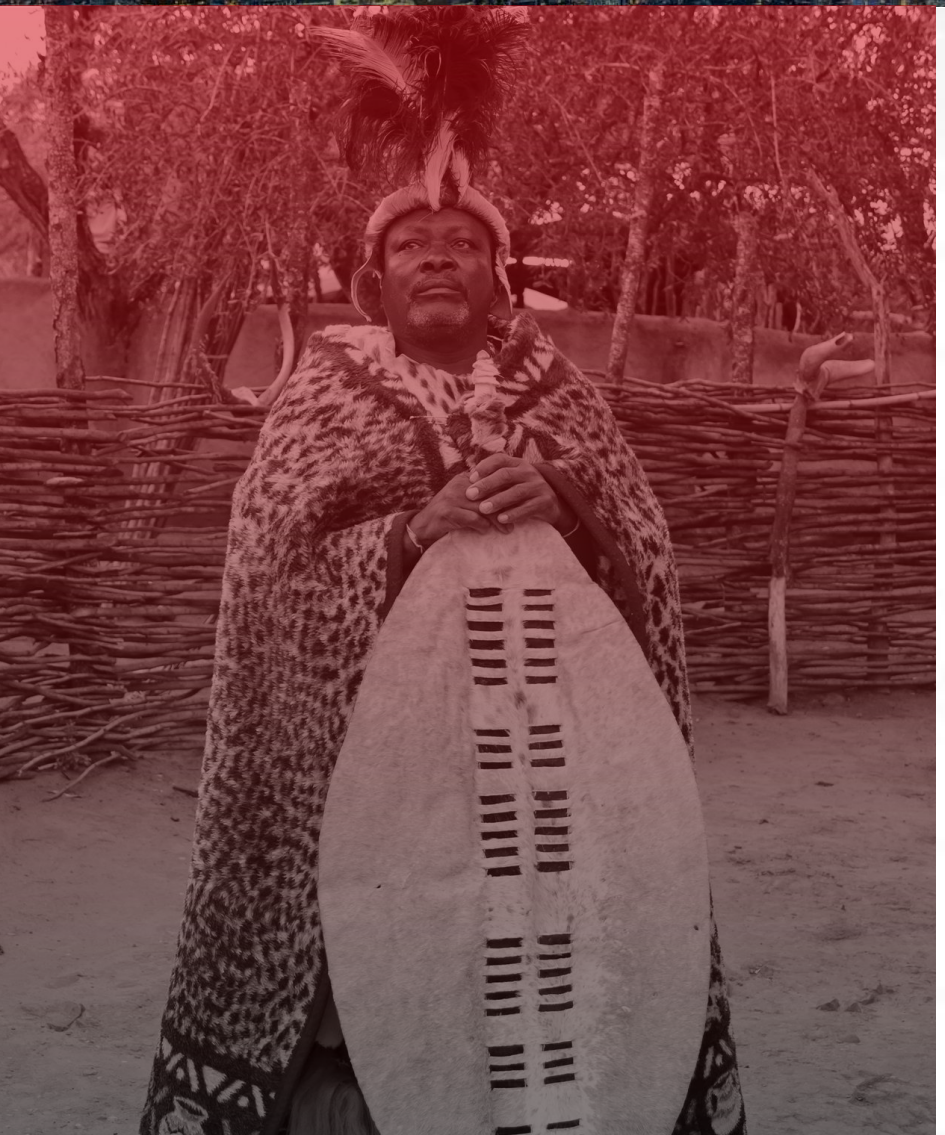
K-Means clustering

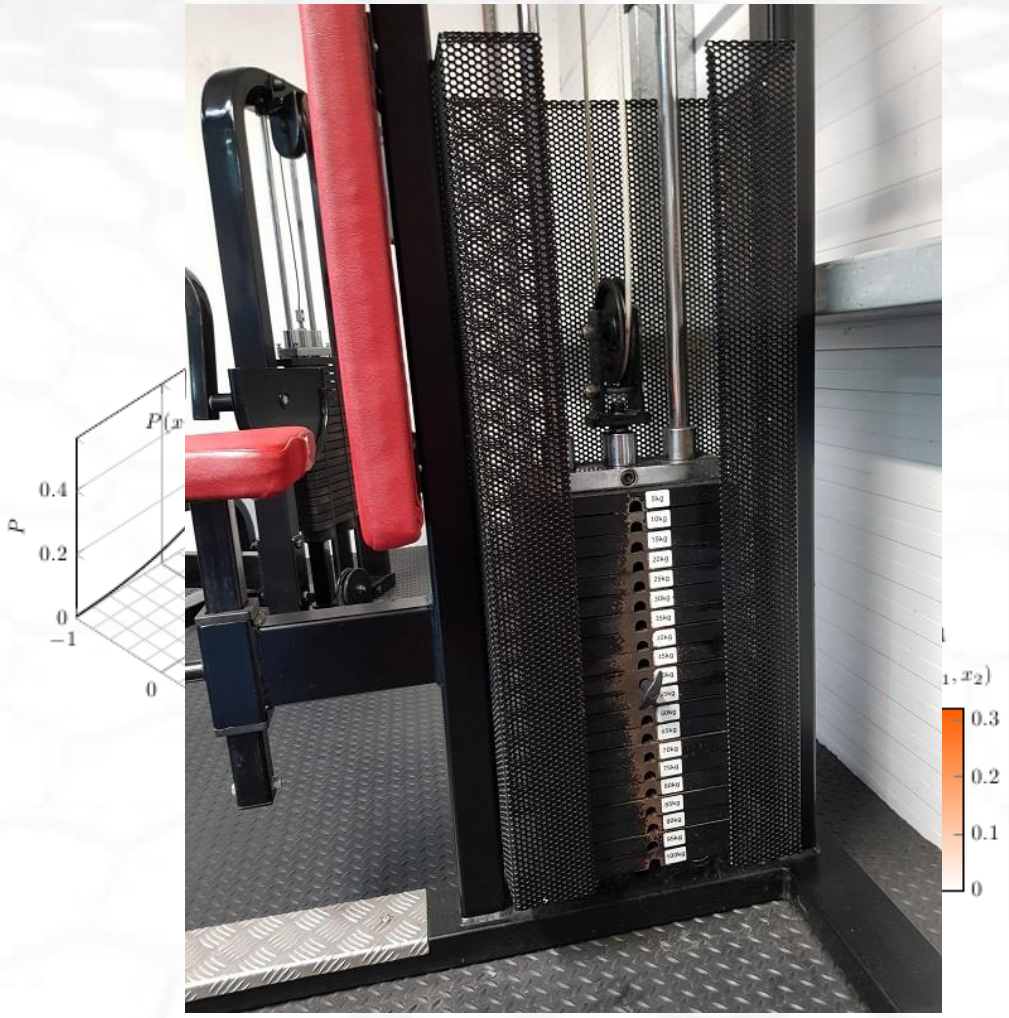
- Decide on number of clusters
- Choose 4 random centres
- Assign data to closest centre
- Move centre chosen to the middle of data assigned
- Assign data to closest centre
- Stop when centres stop moving and number of points assigned to centres don't change

OWASP
AppSec Europe
London 2nd-6th June 2018

Seconds out!

Etienne Greeff
@etienne_greeff



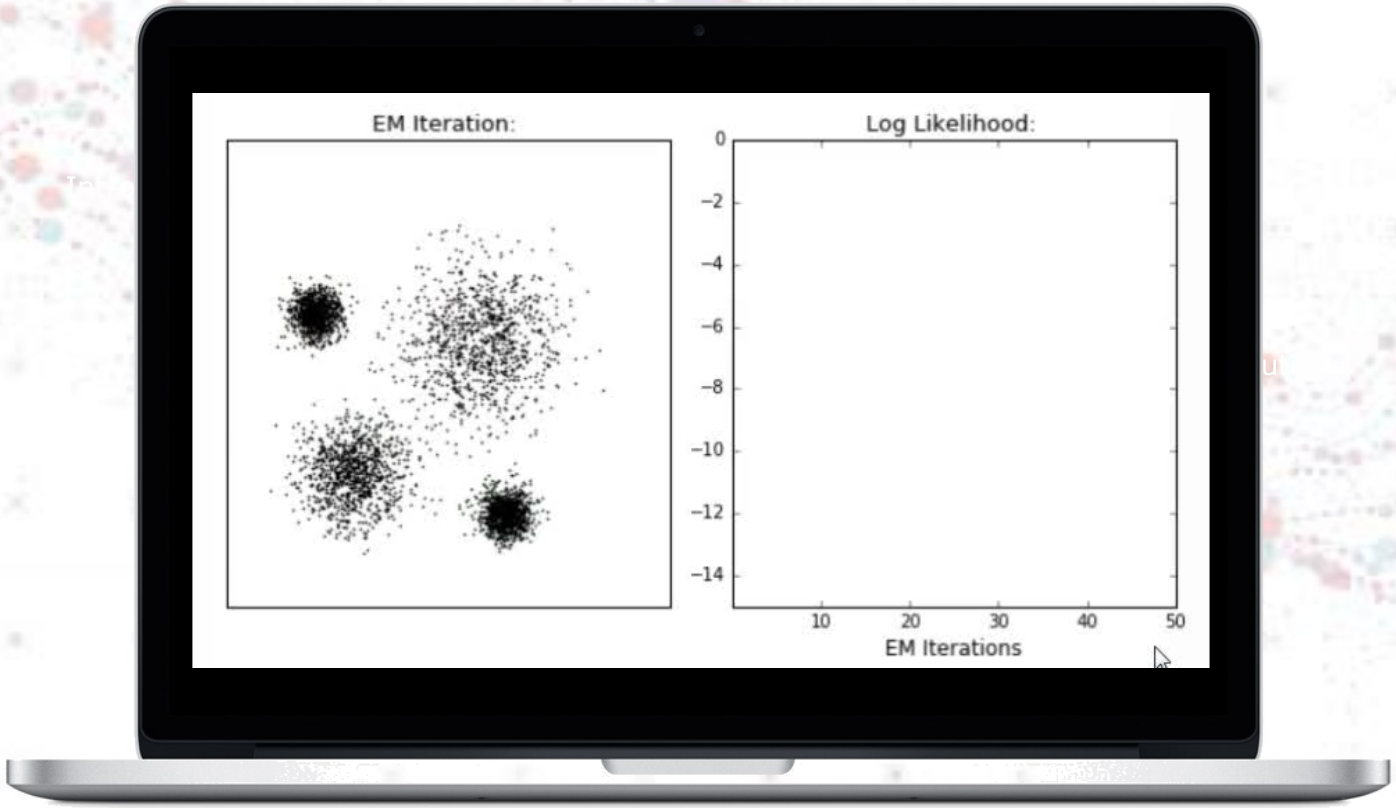


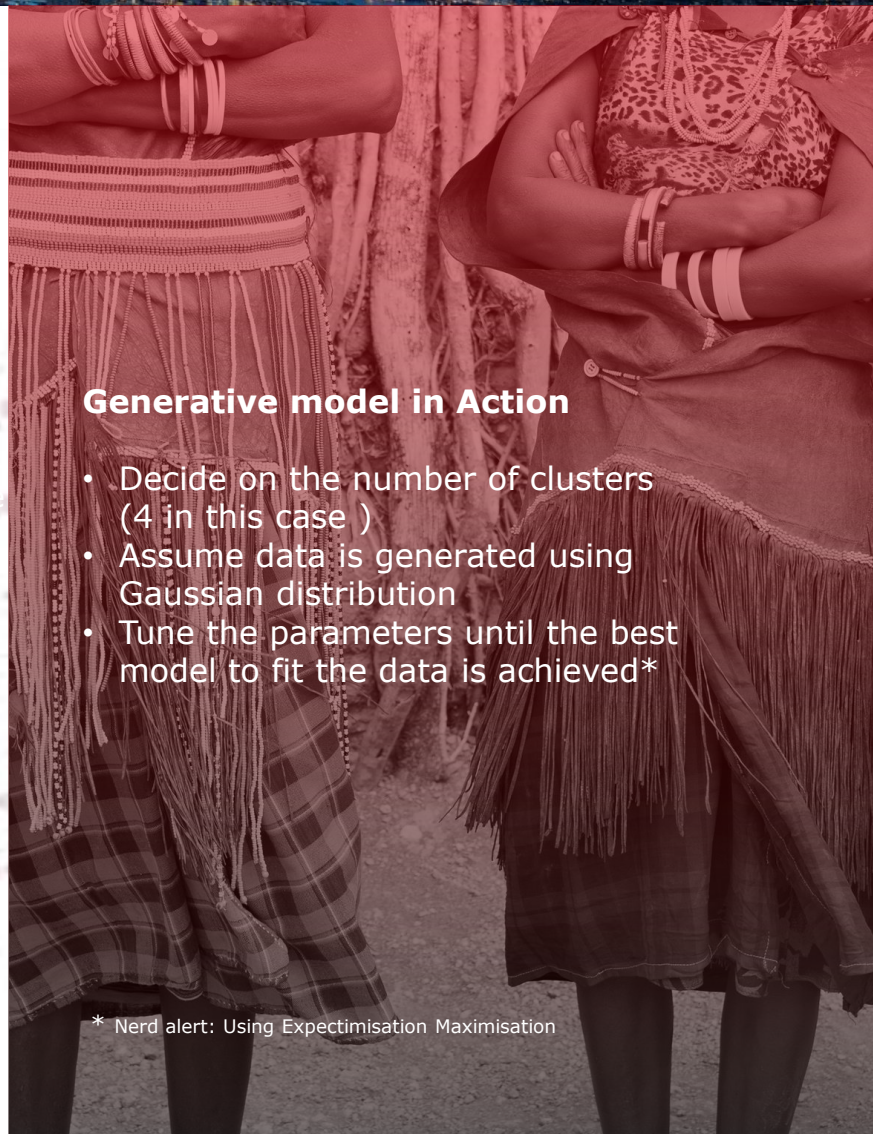
Seconds out!

Etienne Greeff
@etienne_greeff

UNSUPERVISED LEARNING

Using Generative model





Generative model in Action

- Decide on the number of clusters (4 in this case)
- Assume data is generated using Gaussian distribution
- Tune the parameters until the best model to fit the data is achieved*

* Nerd alert: Using Expectimisation Maximisation

OWASP
AppSec Europe
London 2nd-6th June 2018

Seconds out!

Etienne Greeff
@etienne_greeff

SO WHO IS THIS BAYES GUY THAT IS GOING TO SOLVE ALL OUR PROBLEMS?

Thomas Bayes (1702-1761)

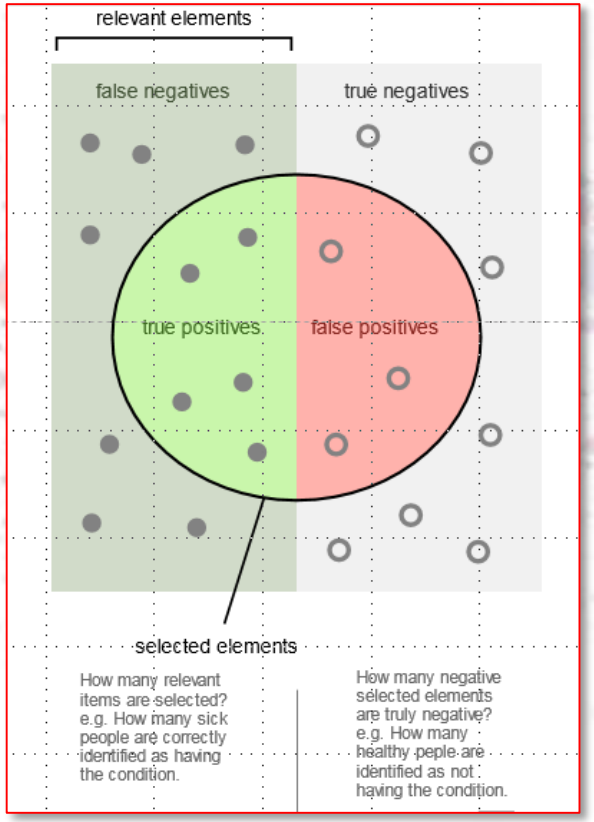
Charlie Sheen (1965-present)
The Movie:
"Return of Thomas Bayes"

My new belief $\propto$ What I am actually seeing x What I believed before

Seconds out!

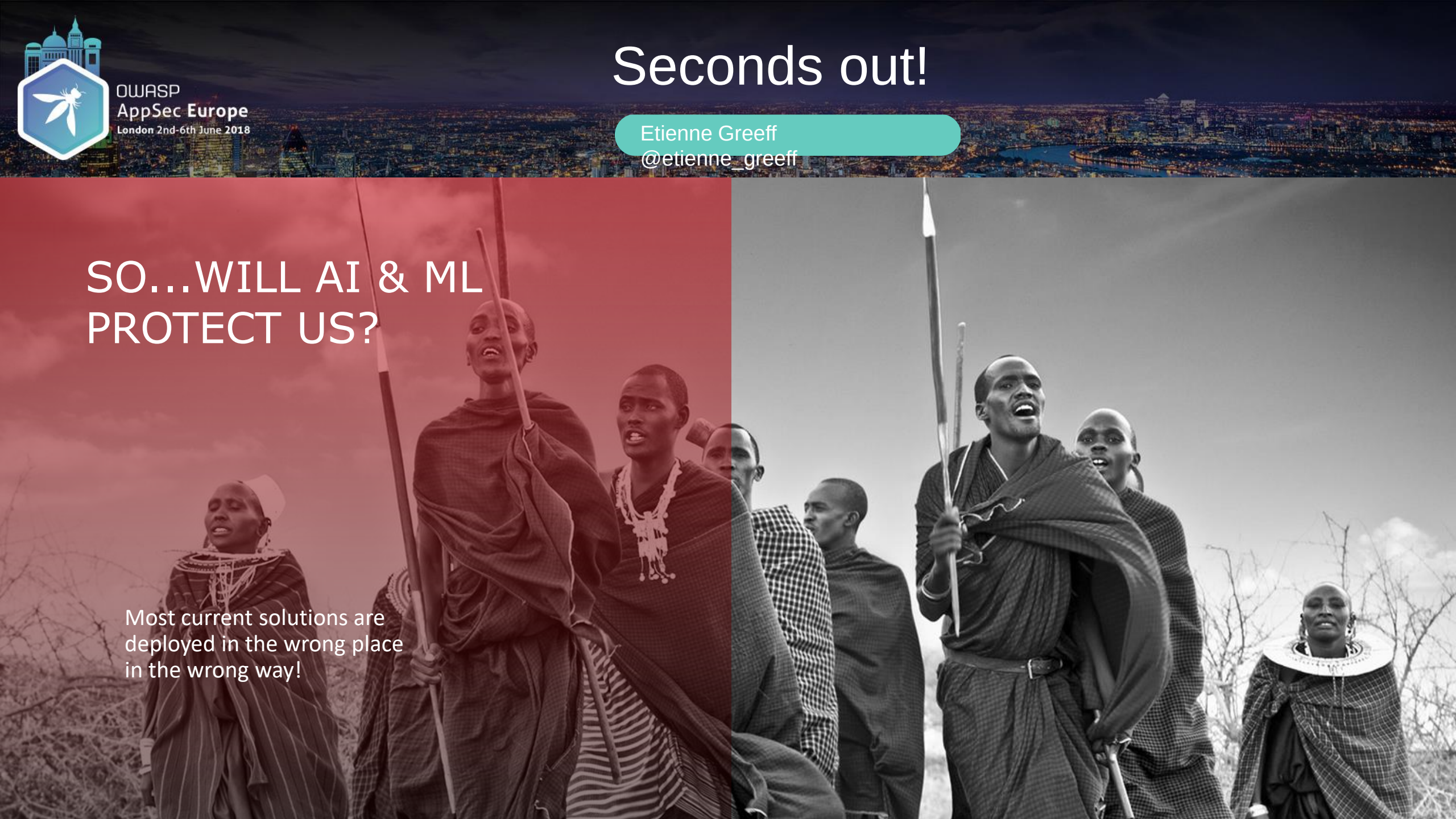
Etienne Greeff
@etienne_greeff

SO, HOW GOOD IS IT?



Sensitivity: (also called the **true positive rate**) measures the proportion of actual positives that are correctly identified as such (e.g., the percentage of anomalous log entries which are correctly identified as anomalous).

Specificity: (also called the **true negative rate**) measures the proportion of actual negatives that are correctly identified as such (e.g., the percentage of log entries which are correctly identified as not being anomalous).



Seconds out!

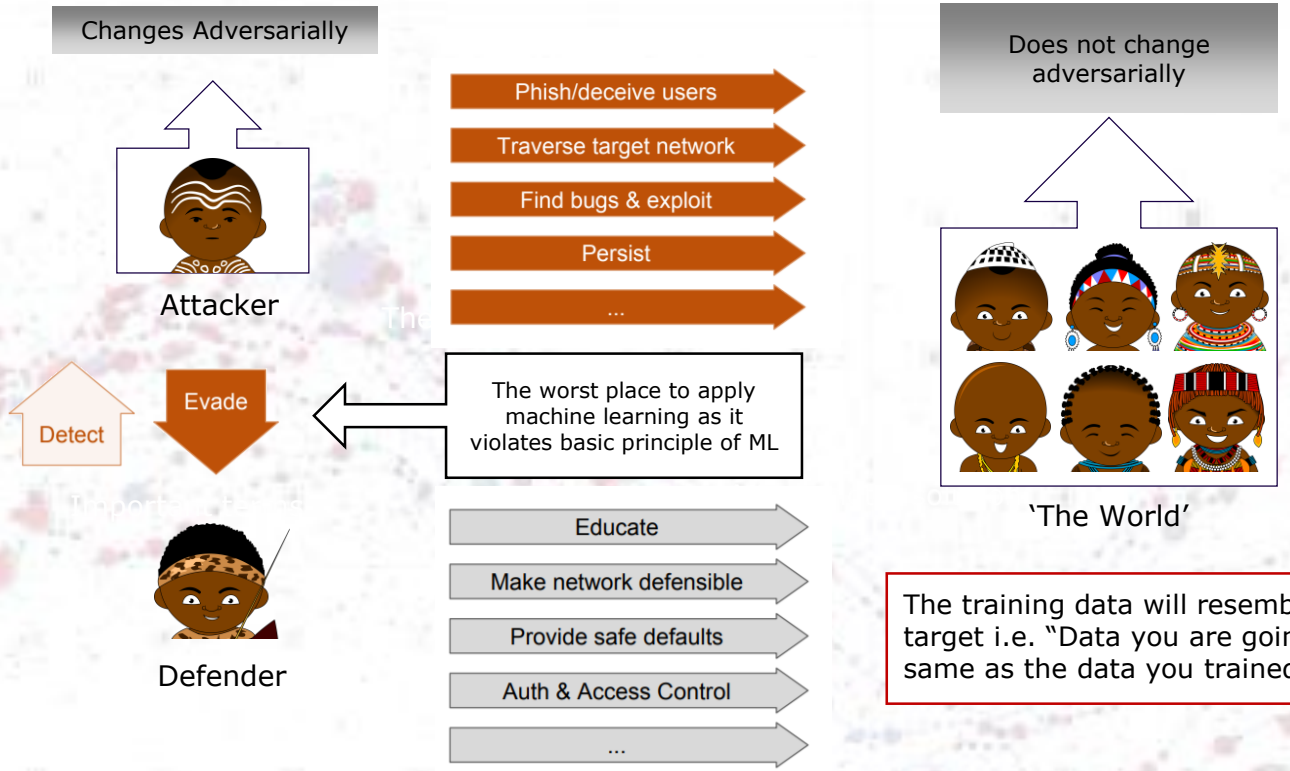
Etienne Greeff
@etienne_greeff

SO...WILL AI & ML PROTECT US?

Most current solutions are deployed in the wrong place in the wrong way!

Seconds out!

Etienne Greeff
@etienne_greeff



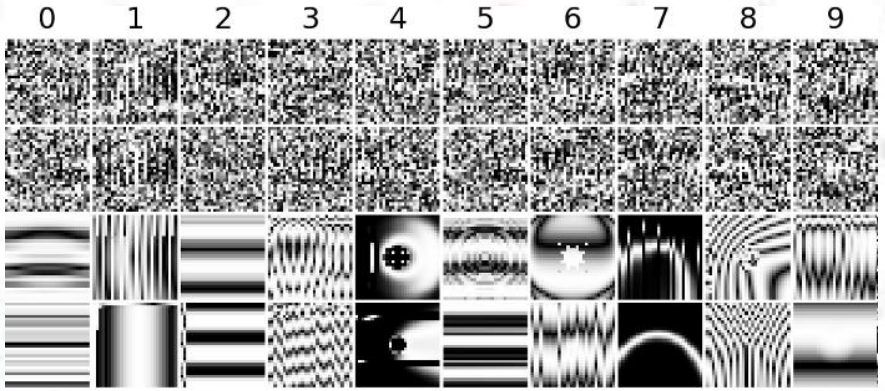
The training data will resemble their real-world deployment target i.e. "Data you are going to work on needs to look the same as the data you trained your ML algorithm on."

With thanks to Thomas Dullien

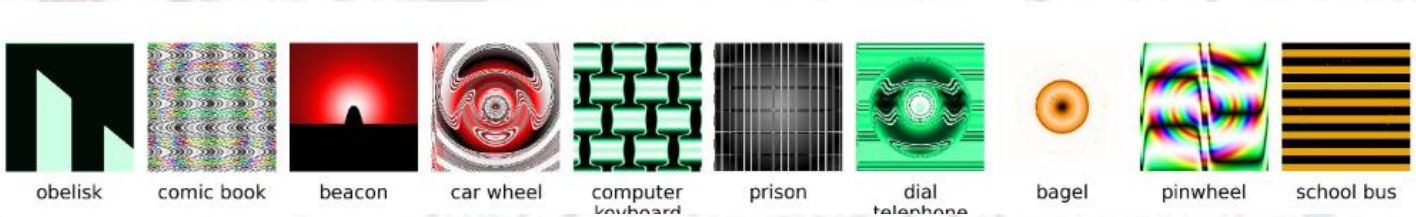
Seconds out!

Etienne Greeff
@etienne_greeff

FOOLING NEURAL NETWORKS FOR FUN & PROFIT



State of the art Neural Network believes with 99.99% probability these represents number 0-9.



Above exploits the structure of how Neural Networks work for instance knowing there are yellow and black edges on a school bus.

Seconds out!

Etienne Greeff
@etienne_greeff

WHY IS THE CURRENT STATE OF PLAY FLAWED?

- Glorified anomaly detection
- Does not work for targeted attacks
- Discriminative model needs a lot of training data
- Training data has been shown to be inaccurate and outdated
- Generative models don’t reflect attackers

Seconds out!

Etienne Greeff
@etienne_greeff

AI & ML AS A TOOL

- Behind all successful AI & ML projects there is a clear problem statement

Here are two examples:

- Identify calls to C&C servers from customer networks
- Analysing large volumes of log messages

Seconds out!

Etienne Greeff
@etienne_greeff

EXAMPLE 1:

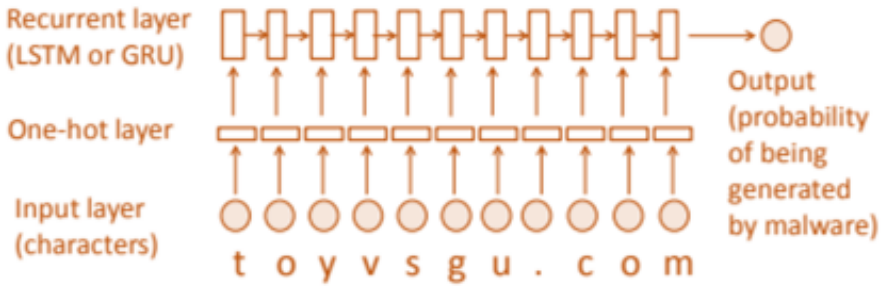
Detecting C&C Traffic

Various families of malware use domain generation algorithms (DGAs) to generate a large number of pseudo-random domain names to connect to a command and control (C2) server.

Seconds out!

Etienne Greeff
@Etienne_greeff

DETECTING C&C TRAFFIC USING NEURAL NETWORK TO EVALUATE DNS NAMES




```
model = Sequential()
model.add(Embedding(max_features, 128, input_length=maxlen))
model.add(LSTM(128))
model.add(Dropout(0.5))
model.add(Dense(1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy',
              optimizer='rmsprop')
```

Training dataset:

- Equal number of benign and malicious domains
- Malicious domains from banjori, corebot, cryptolocker, dircrypt, kraken, lockyv2, pykspa, qakbot, ramdo, ramnit, simda

BANJORI.

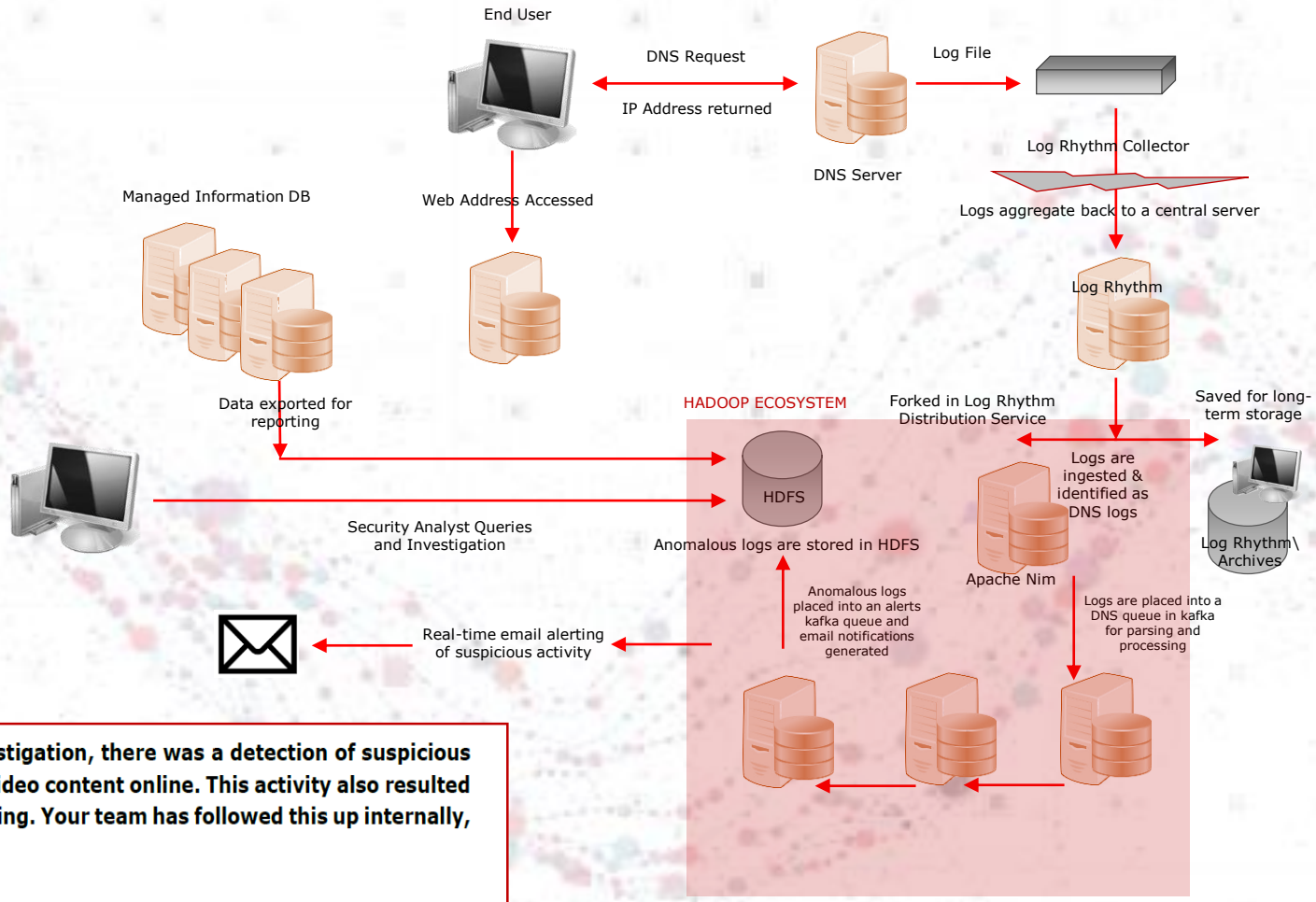
Cryptolocker

coreboot

Seconds out!

Etienne Greeff
@etienne_greeff

DETECTING C&C PRODUCTION SETUP



As part of our Domain Name Generated Algorithm routine investigation, there was a detection of suspicious activity from a user which transpired to relate to streaming of video content online. This activity also resulted in potential crypto-coin mining activity because of video streaming. Your team has followed this up internally,

MTD REPORT CUSTOMER IN CONFIDENCE

Seconds out!

Etienne Greeff
@etienne_greeff

DETECTING C&C
WHAT THE TEAM SEES

In this example it has detected the domain 10ak7u9vn1dl01rnuo65k1i3qv.net.

Data	
Alarm ID	2348200
Alarm Date	03/16/2018 3:17:55 pm
Alarm Name	AIE: SD: Domain Generated Algorithm Detected
Alarm Description	AIE: Domain generation algorithms (DGA) are algorithms seen in various families of malware that are used to periodically generate a large number of domain names that can be used as rendezvous points with their command and control servers.
Classification	Suspicious
Log Source	AI Engine (AIEEngineID: 5) (-1000510)
Common Event	AIE: SD: Domain Generated Algorithm Detected
Direction	Unknown
Entity (Origin)	MTD Services
Entity (Impacted)	MTD Services
Host (Origin)	192.168.18.35
User (Impacted)	securedata
Domain (Impacted)	10ak7u9vn1dl01rnuo65k1i3qv.net
Severity	0.9991069436073303

Seconds out!

Etienne Greeff
@etienne_greeff

EXAMPLE 2: ANALYSING LARGE VOLUMES OF DATA

Our analysts need to review a very large number of log messages from endpoint applications and operating systems on a daily basis focusing on the messages that matter*

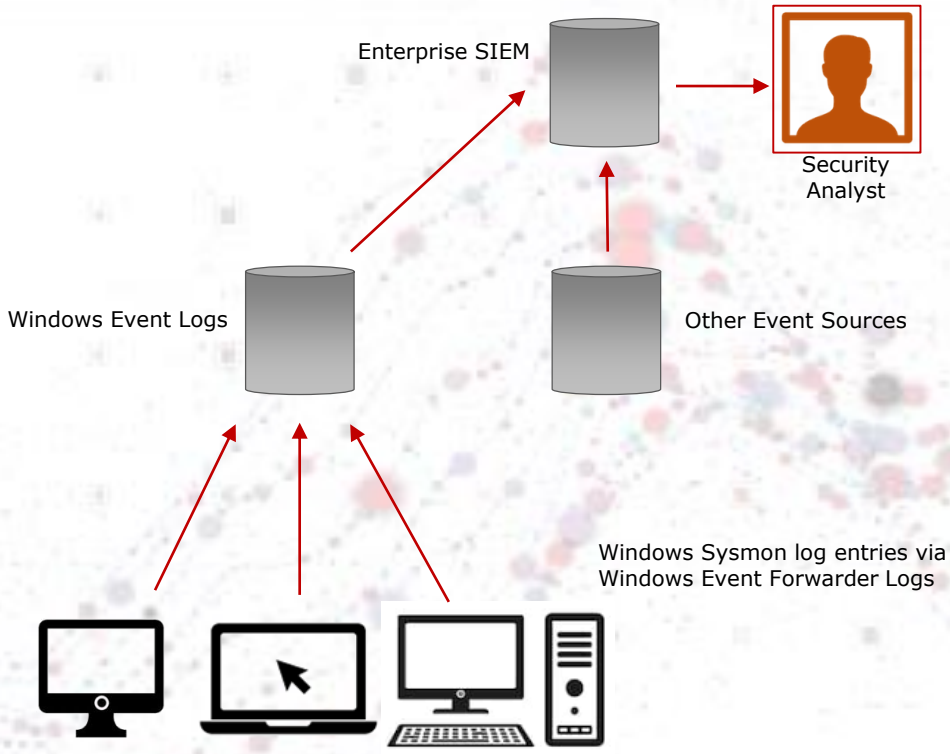
*the system needs the ability to become smarter as we learn more

Seconds out!

Etienne Greeff
@etienne_greeff

DEALING WITH 25,000 LOG ENTRIES PER DAY

- We use sysmon monitoring on endpoints to log pertinent events
- Typically will receive about 25,000 entries per customer, per day
- It's not possible to go through all the entries
- Would like to group similar entries together so we can analyse quickly
- Would be good if the system can get smarter over time as we identify both good, interesting and obviously malicious entries



Seconds out!

Etienne Greeff
@etienne_greeff

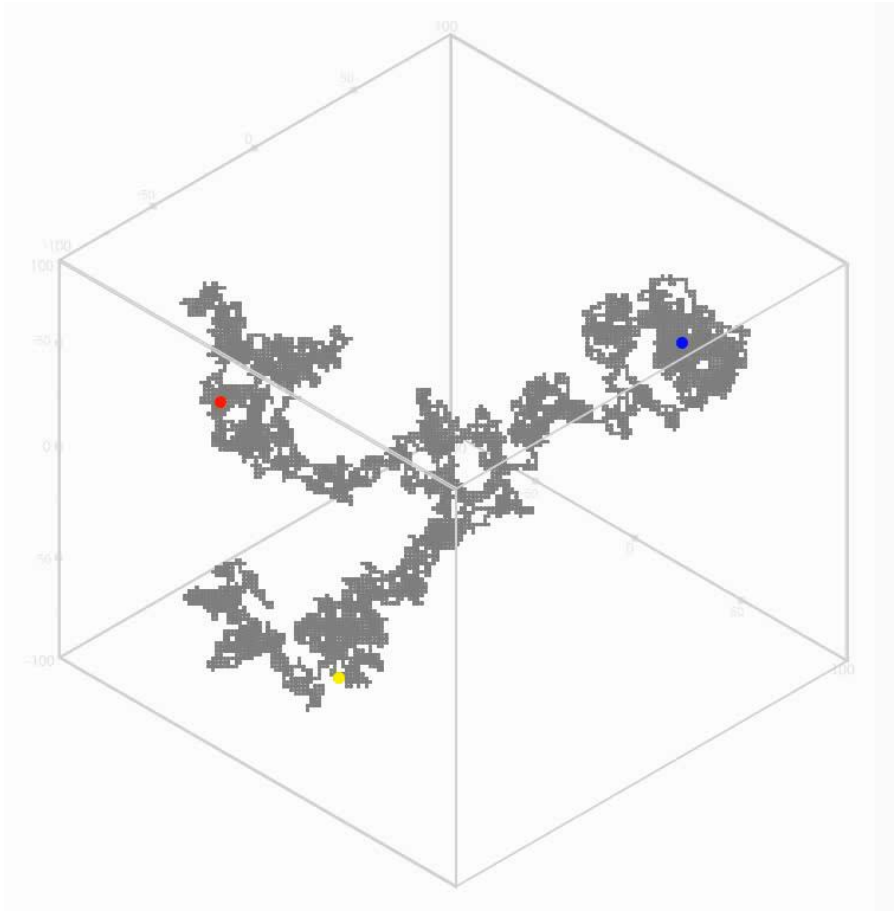
THE RULES

Label ID	Label	Comment
0	Apparently benign script execution	Evoking a script from cmd.exe is fundamentally suspiicious. In this c
1	Innocent DLL registration	Regsvr is being used to register a know safe G2M DLL
2	Apparently benign script execution	Cmder is a known terminal emulator for Windows
3	Boring	This is a predictable combination of Explorer and Office clicktorun
4	Suspicious action by a known executable	Regsv is being evoked, but by an exe that we recognise
5	Suspicious action by a known executable	Wscript is being evoked, but most likely by Skype for Business
6	Boring	Looks like a common sequence
7	Suspicious action by a known executable	This is Okta - a known and expected application
8	Interesting Command Line	Viscocity is a known VPN client, but anything evoking ipconfig is wor
9	Interesting Command Line	Netstat being executed by system. Just seems odd.
10	Apparently benign script execution	The combination of docker and Powershell is probably normal, but i
11	Interesting Command Line	Looks benign, but netwok utilities seem worth watching
12	Suspicious action by a known executable	This is just unusual, though it would need further investigation ot m
13	Suspicious script execution	Any use of Powershell that's not obviously benign is worth looking i
14	Executable run from a suspocious location	Running an exe from a temp directory is worth looking into
15	Probably benign browser download to temp	Probably benign browser download to temp
16	Probably benign app dump to temp file	Error reporter writing to a dump file
17	Probably benign app dump to temp file	Java writing some kind of state
18	Suspicious script execution	Powershell. Temp location. Random name.

Seconds out!

Etienne Greeff
@etienne_greeff

MARKOV CHAIN BASED RANDOM WALK SEMI-SUPERVISED CLASSIFIER



OWASP
AppSec Europe
London 2nd-6th June 2018

Seconds out!

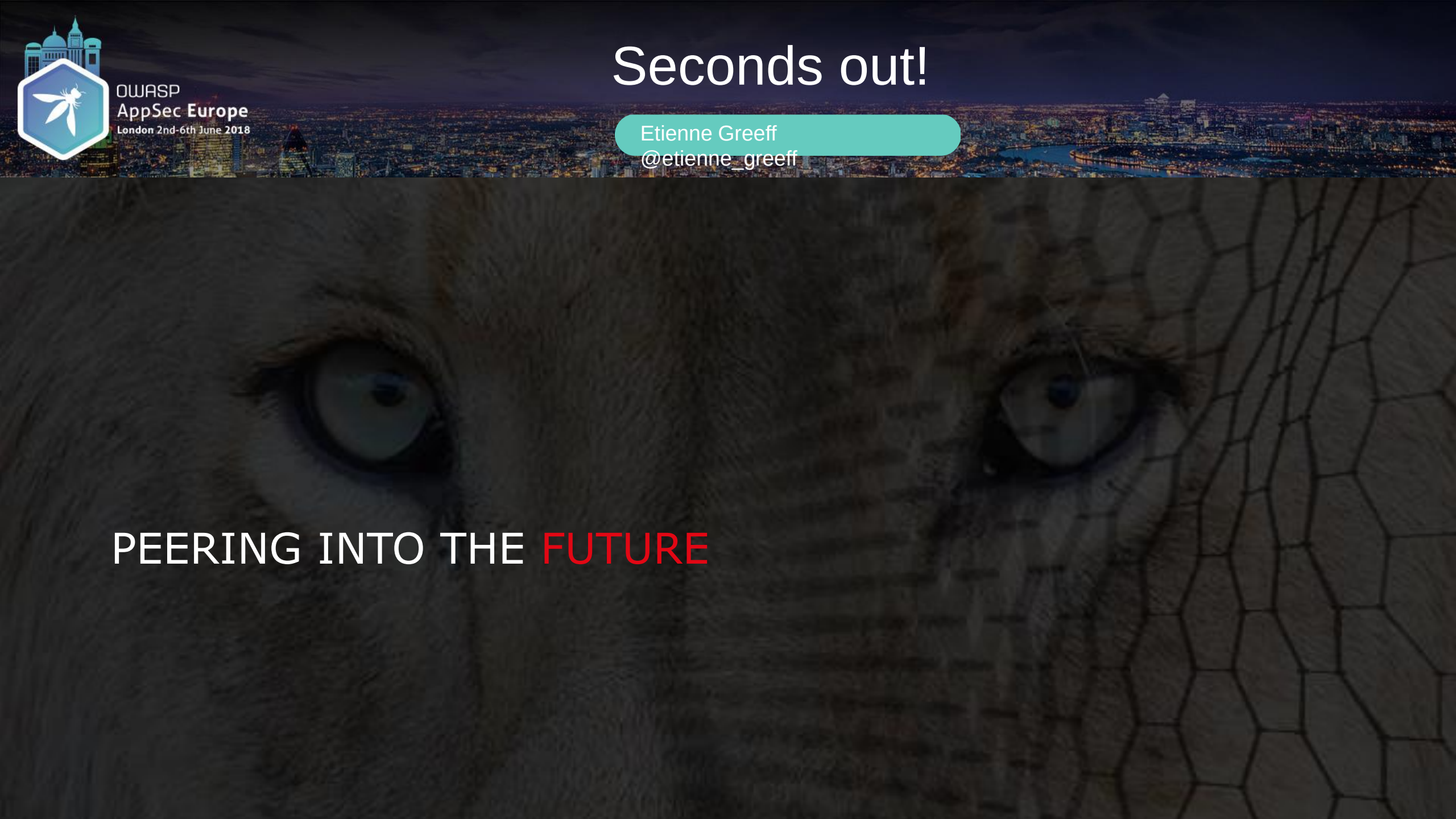
Etienne Greeff
@etienne_greeff

THE RESULTS

Rule Name	Labe	Label	Command	Process	Probability
AIE: SD: SysMon: Startup/Temp Locations	21	Probably benign browser download to ten	NoCommand	c:\windows\systemapps\microsoft.microsoftedge_8wekyb3d8bbwe\microsoftedge.ex	0.3383
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743
AIE: SD: SysMon: Browser Launched By Office Ap	12	Suspicious action by a known executable	c:\program files\microsoft office\root\office16\lync.exe/fromrunkey	c:\program files (x86)\google\chrome\application\chrome.exe	0.34743

Identifies log entries that are similar to other rules but without explicit rules i.e. identifies behaviour of Microsoft Edge launch by Lync as same as Chrome launched by Lync although no explicit rule!

Having to analyse 100 entries manually now rather than 25,000


OWASP
AppSec Europe
London 2nd-6th June 2018
Seconds out!
Etienne Greeff
@etienne_greeff
PEERING INTO THE FUTURE

Seconds out!

Etienne Greeff
@etienne_greeff

CYBER SECURITY TODAY Vs AI ENABLED THREAT OF TOMORROW

Cybersecurity	AI enabled threat of the future
Typo squatting, phono squatting	Voice squatting and voice masquerading
Penetration testing & Fuzzing applications for fun and profit	Reinforcement learning based fuzzing and machine learning based vulnerability discovery
Feature based attacks on the new battleground i.e. the endpoint	Using rich machine learning features on the endpoint

Seconds out!

Etienne Greeff
@etienne_greeff

TYPO SQUATTING

1 Select an industry

Dubbed 'Friday Afternoon Fraud', the conveyancing scam has been known to take several forms, but generally occurs when the hackers intercept emails between home buyers or sellers, and their solicitors.

They generate lookalike emails which allow them to pose as the solicitor involved.
During the final stages of a property purchase or sale, they inform potential victims by email that certain bank account details have changed.

Home buyers stand to lose thousands in new cyberattack

Cybercriminals are hacking the email accounts of Irish solicitors in an attempt to steal tens of thousands of euro from unsuspecting home buyers, the Sunday Independent has learned. Stock photo: PA

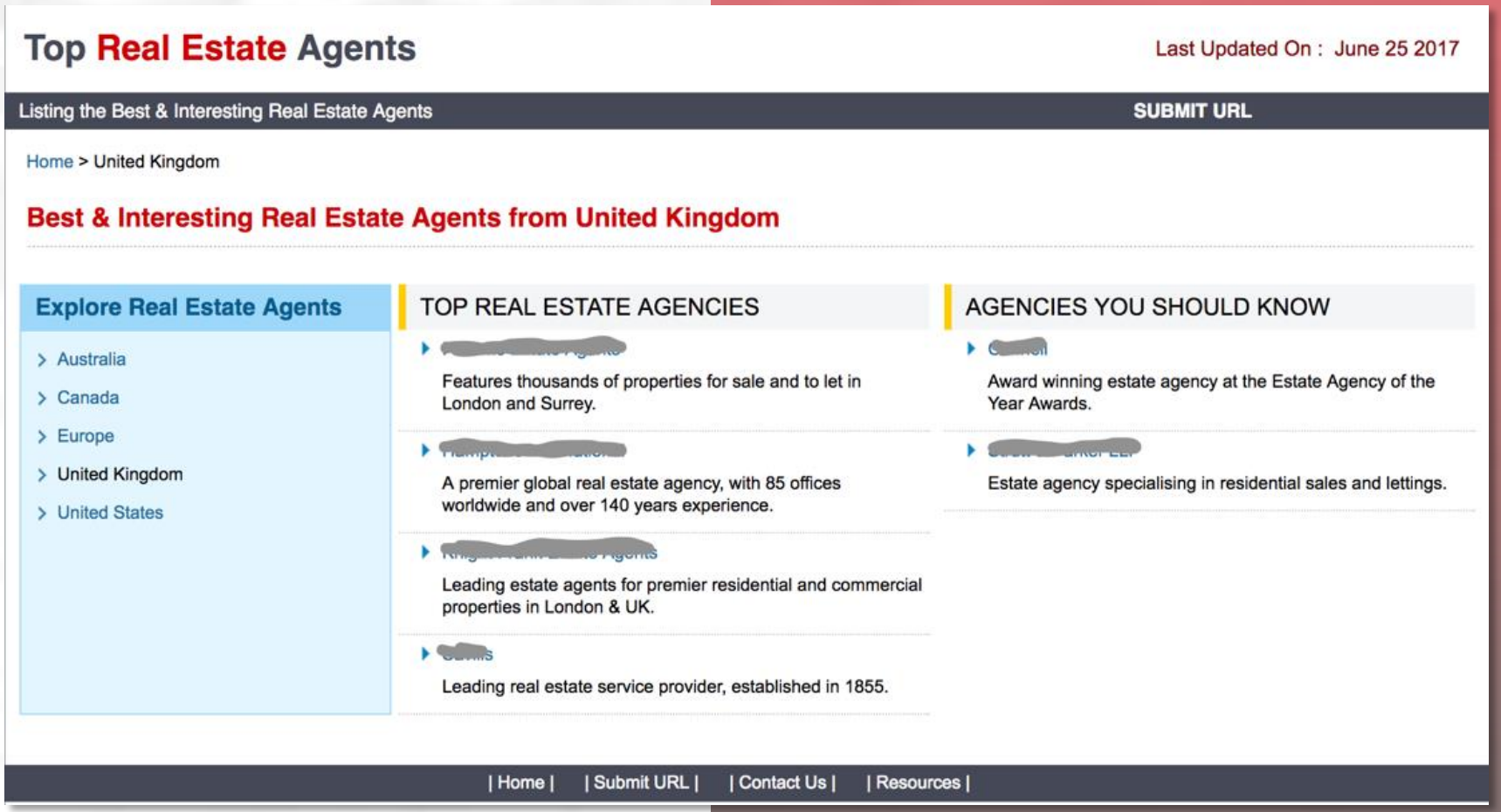
Mark O'Regan
February 5 2017 2:30 AM

Seconds out!

Etienne Greeff
@etienne_greeff

TYPO SQUATTING

2 Enumerate the players

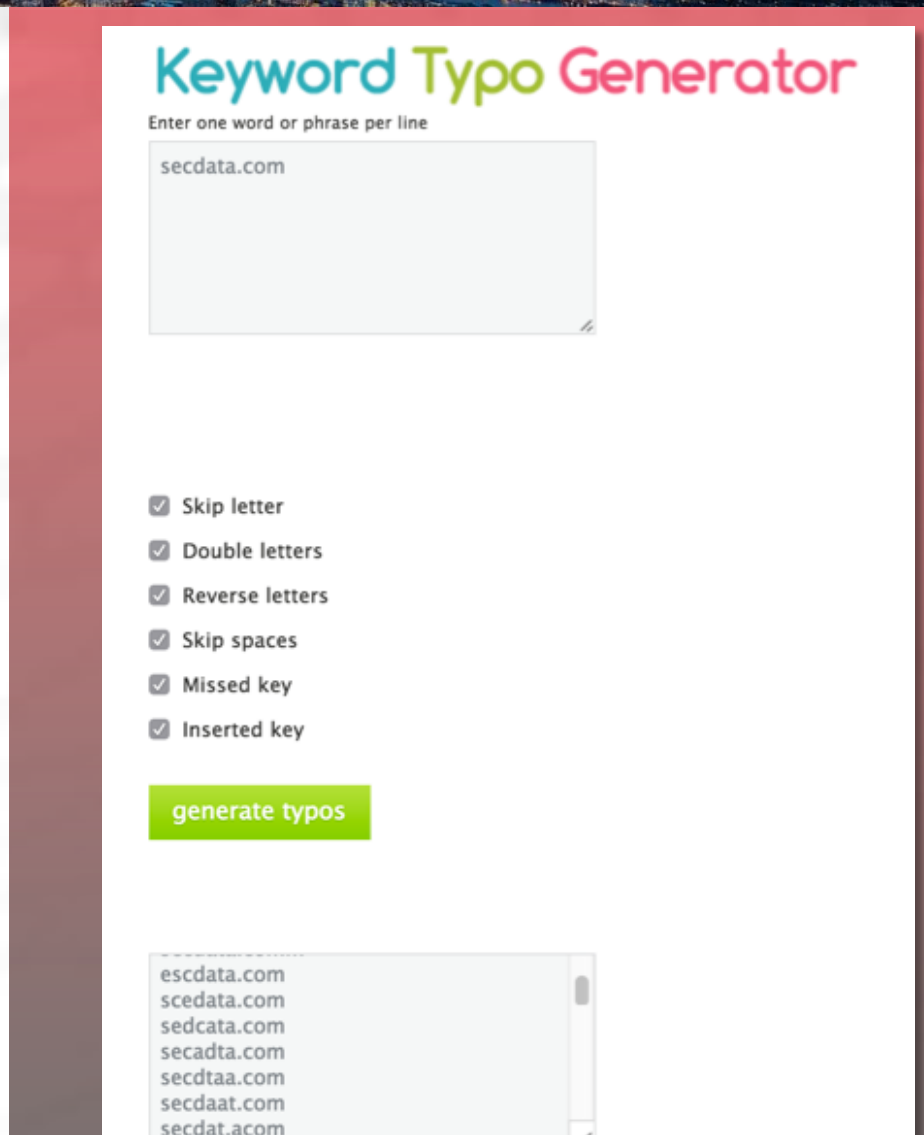


Seconds out!

Etienne Greeff
@etienne_greeff

TYPO SQUATTING

3 Generate the typos

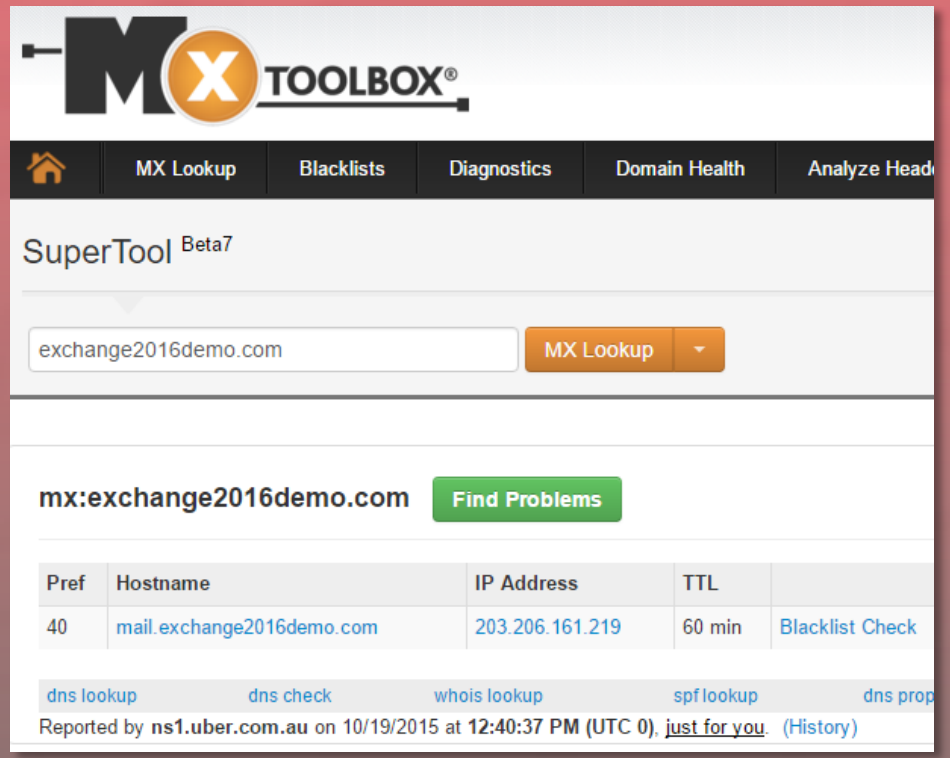


Seconds out!

Etienne Greeff
@etienne_greeff

TYPO SQUATTING

4 Register the domains & setup mail server & wait just wait

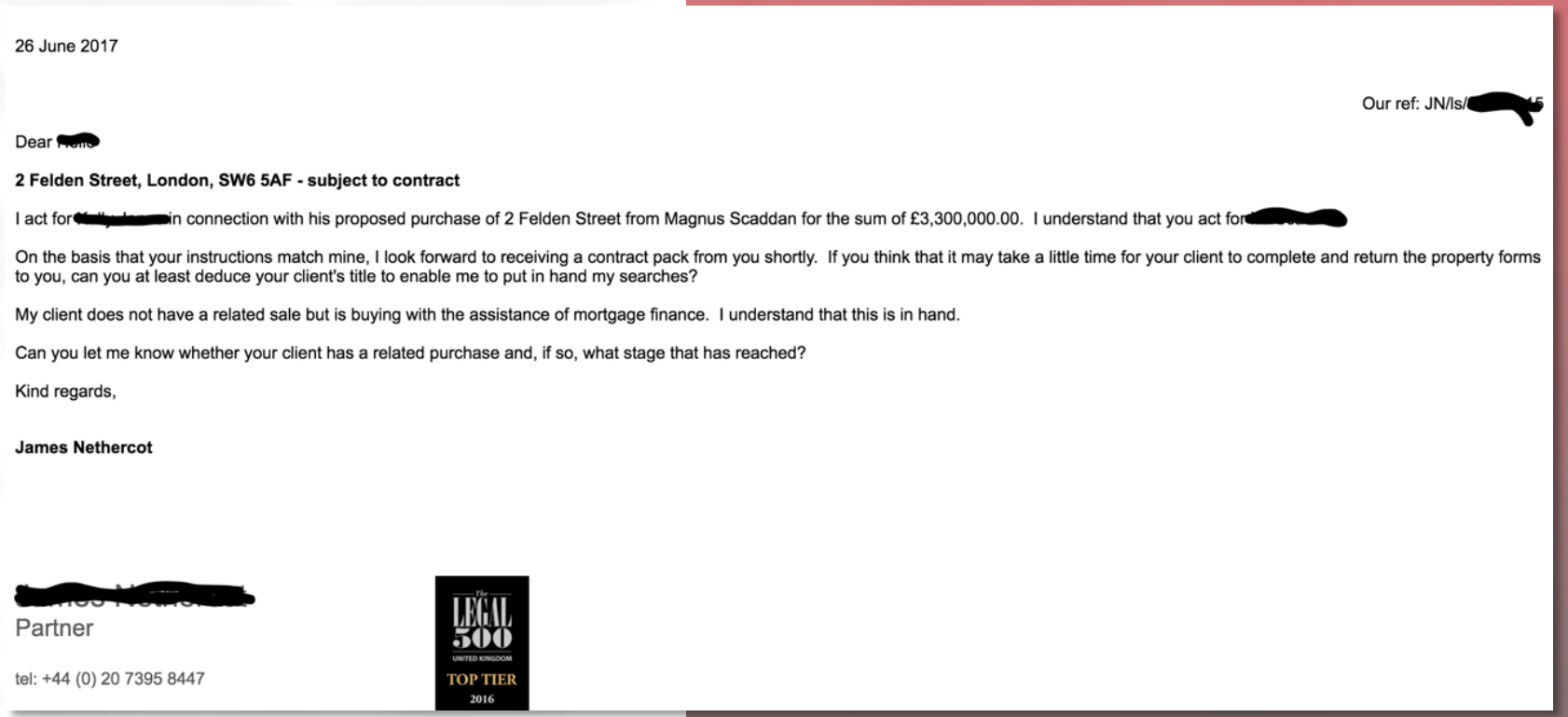


Seconds out!

Etienne Greeff
@etienne_greeff

TYPO SQUATTING

5 Great success!



26 June 2017

Our ref: JN/ls/[redacted]

Dear [redacted]

2 Felden Street, London, SW6 5AF - subject to contract

I act for [redacted] in connection with his proposed purchase of 2 Felden Street from Magnus Scaddan for the sum of £3,300,000.00. I understand that you act for [redacted]

On the basis that your instructions match mine, I look forward to receiving a contract pack from you shortly. If you think that it may take a little time for your client to complete and return the property forms to you, can you at least deduce your client's title to enable me to put in hand my searches?

My client does not have a related sale but is buying with the assistance of mortgage finance. I understand that this is in hand.

Can you let me know whether your client has a related purchase and, if so, what stage that has reached?

Kind regards,

James Nethercot

[redacted]
Partner

tel: +44 (0) 20 7395 8447

The LEGAL 500 UNITED KINGDOM TOP TIER 2016

Seconds out!

Etienne Greeff
@etienne_greeff

AI EQUIVALENT OF TYPO SQUATTING

A.k.a. Voice Squatting and Voice Masquerading

Seconds out!

Etienne Greeff
@etienne_greeff

WHEN ENQUIRING ABOUT CATS GETS CONFUSING

Cats on the Alexa market place

- 66 different Alexa skills are called cat facts
- 5 called cat fact
- 11 whose invocation names contain the string “cat fact”, e.g. fun cat facts, funny cat facts

Seconds out!

Etienne Greeff
@etienne_greeff

HERE IS A REAL WORLD EXAMPLE: WHEN BEING POLITE COSTS YOU...

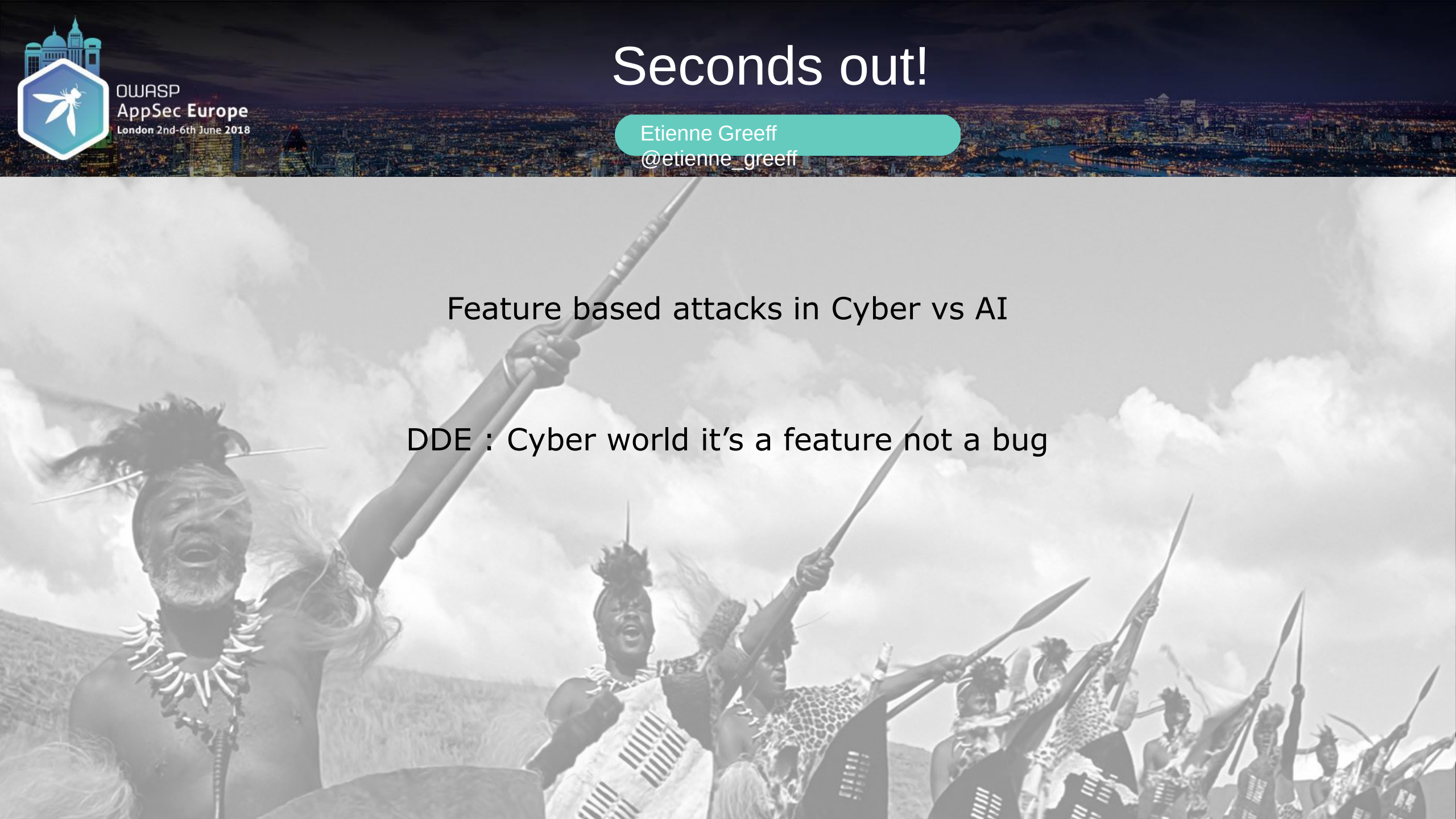
Voice Squatting:

- Adding a malicious skill to market place that impersonates another skill i.e. Captitol One or Please Capital One or Capital Won instead of valid skill Capital One
- At present, the system is more likely to match malicious skill invoked by "*Please Open Capital One*" than proper skill behind "*Open Capital One*"
- 51% of people use polite words before skills i.e. please can you…

Voice Masquerading

- When the trust system on the VPA is abused
- The VPA relies on the current running skill (which may be malicious) to stop
- The malicious skill is then in a position to gather all types of confidential information as it continues running
- Real life tests show this is possible and people don't pay attention to light indicating skill is active

There is evidence of multiple skills that could be abused in the above ways



Seconds out!

Etienne Greeff
@etienne_greeff

Feature based attacks in Cyber vs AI

DDE : Cyber world it’s a feature not a bug

OWASP
AppSec Europe
London 2nd-6th June 2018

Seconds out!

Etienne Greeff
@etienne_greeff

mutt | saif@5A1f151-1E13E: ~/research/tmp | sudo ncat -lvp 443

sudo ncat -lvp 443 116x31

```
┌─[~/research/tmp]─[saif@5A1f151-1E13E]─[0]─[4805]
└─[:)] % sudo ncat -lvp 443
Ncat: Version 7.12 ( https://nmap.org/ncat )
Ncat: Listening on :::443
Ncat: Listening on 0.0.0.0:443
```

sudo python -m SimpleHTTPServer 80 116x31

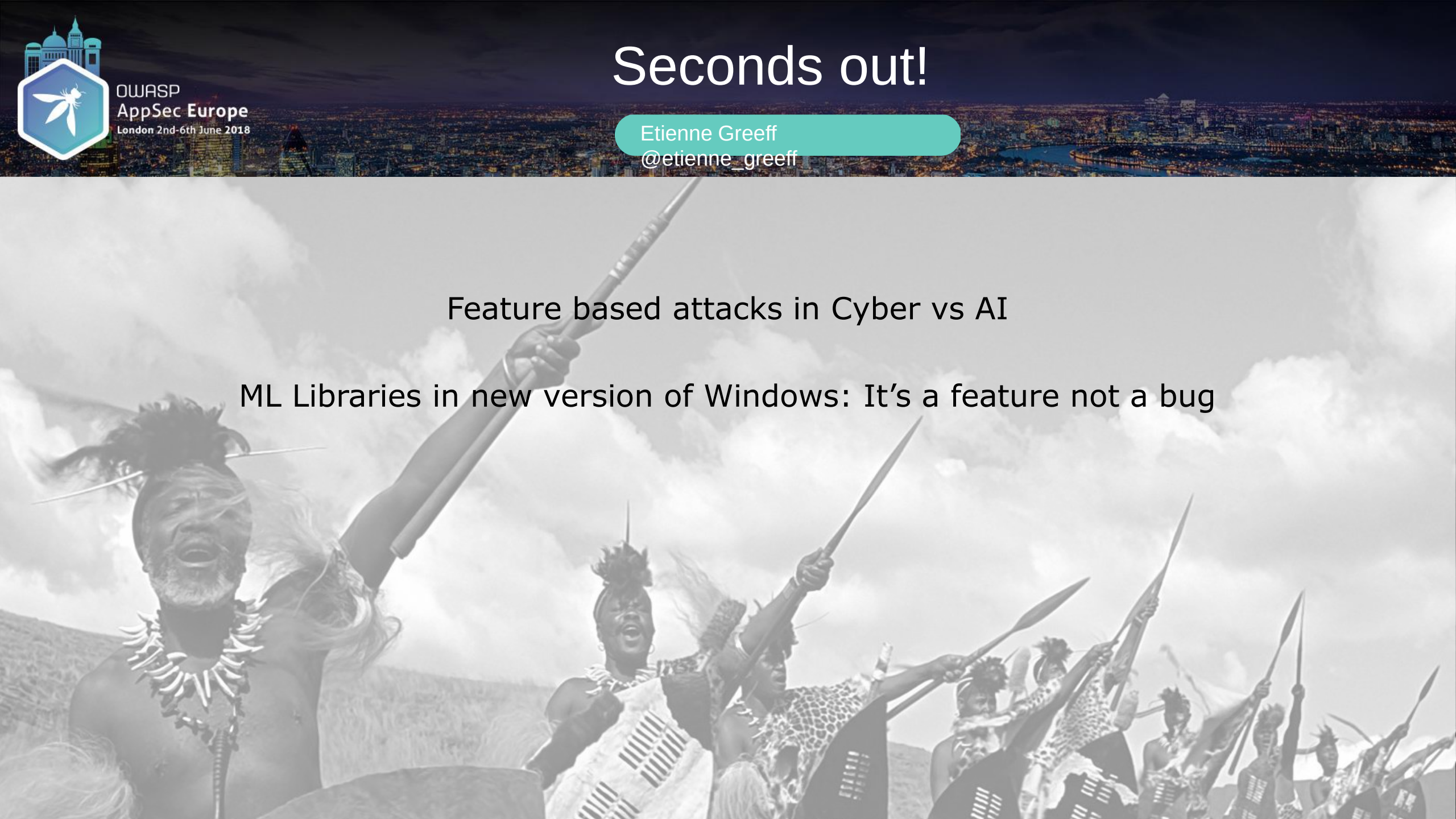
```
┌─[~/research/tmp]─[saif@5A1f151-1E13E]─[S]─[4790]
└─[ ] % sudo python -m SimpleHTTPServer  80
[sudo] password for saif:
Serving HTTP on 0.0.0.0 port 80 ...
```

My Computer | Windows_8_x64

C:\Windows\system32\cmd.exe

```
C:\Users\sensepsot>_
```

To direct input to this VM, move the mouse pointer inside or press Ctrl+G.



Seconds out!

Etienne Greeff
@etienne_greeff

Feature based attacks in Cyber vs AI

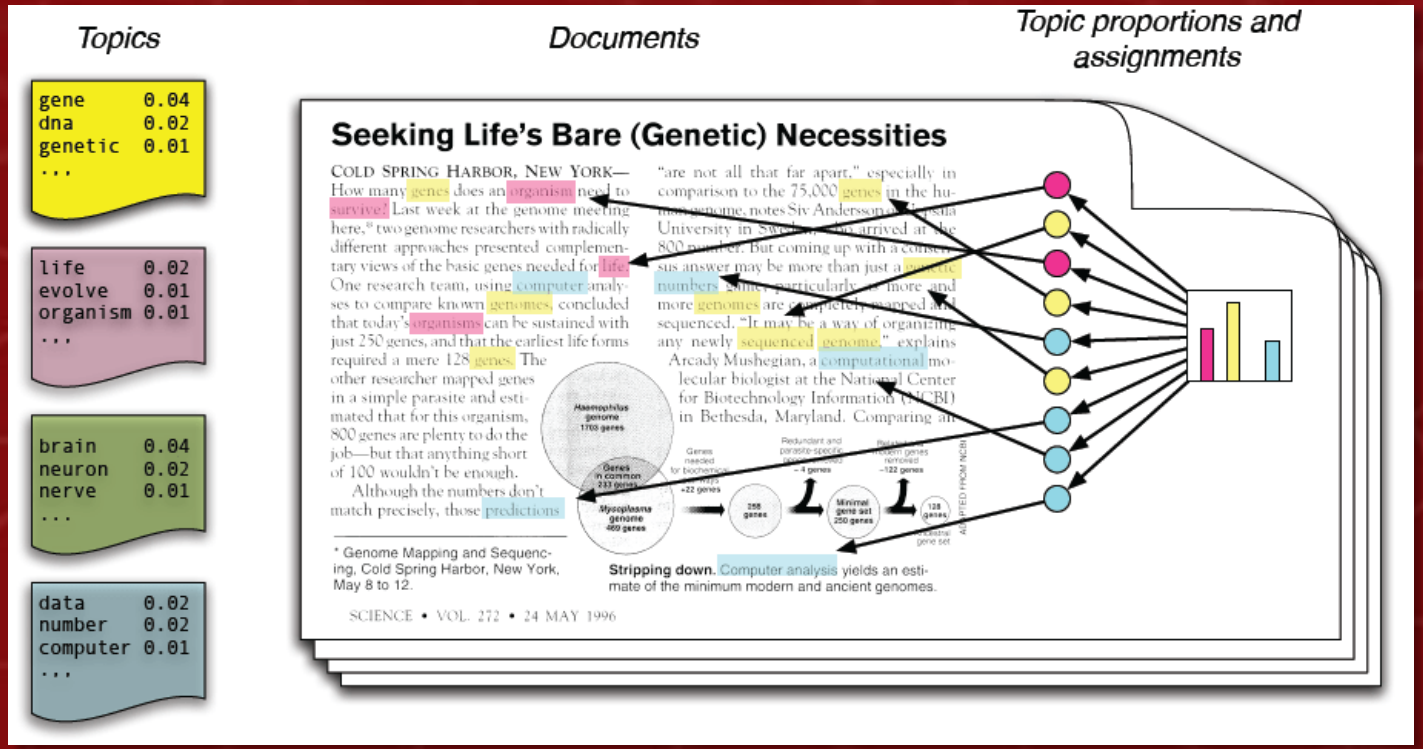
ML Libraries in new version of Windows: It’s a feature not a bug

Seconds out!

Etienne Greeff
@Etienne_greeff

TOPIC MODELLING FOR FUN AND PROFIT ON DESKTOPS

Topic modelling is the process of analysing which words are used together the most in the most common ways, in other words what topics does this bunch of documents discuss using which words.

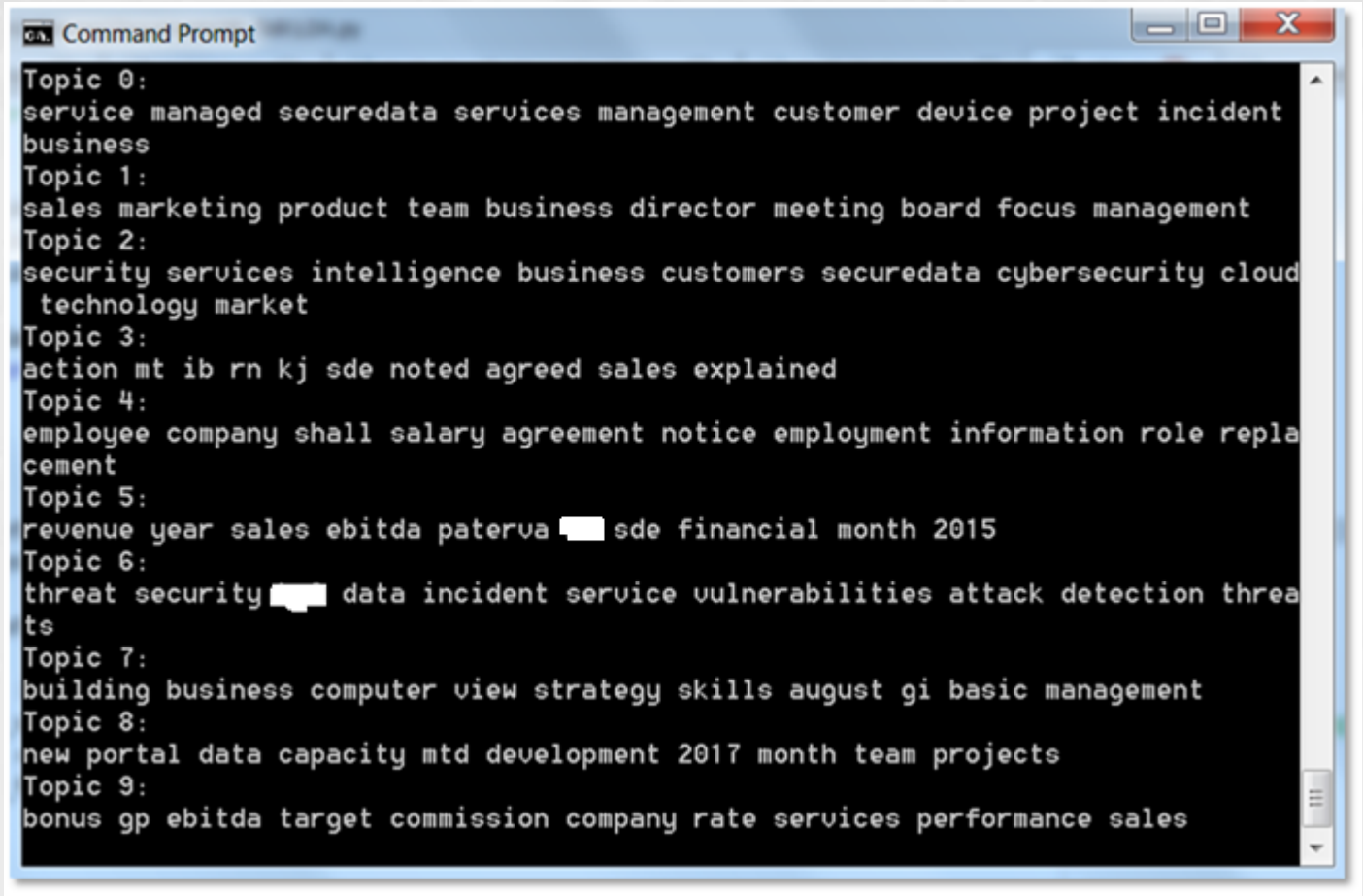


Seconds out!

Etienne Greeff
@etienne_greeff

TOPIC MODELLING FOR FUN AND PROFIT ON DESKTOPS

- My Desktop
- 2 minutes, 800 files
- Scarily accurate
- Unparalleled insight
- Identify valuable information

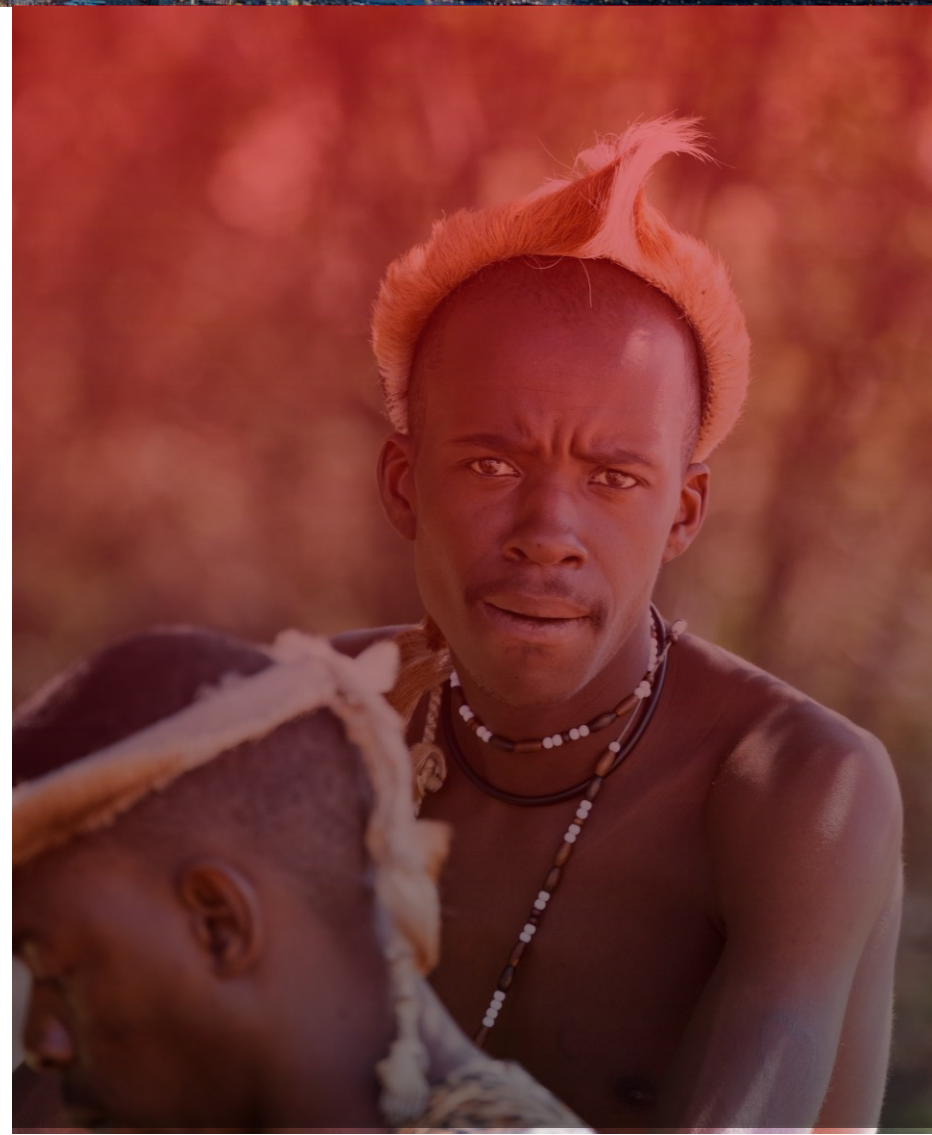


Seconds out!

Etienne Greeff
@etienne_greeff

RESPONSIBLE DISCLOSURE IN THE AI & ML WORLD

- Mostly disclosed in academic papers first
- With software vulnerabilities we all know the rules
- With AI & ML things are a bit different
    - Algorithm weaknesses affect multiple applications including IOT and embedded systems
    - No culture of responsible disclosure
    - AI & ML innovation is mostly in the academic domain at the moment
    - Who do you notify?
- There is a movement to create rules for responsible disclosure



Seconds out!

Etienne Greeff
@etienne_greeff

HOW DO I GET TO GRIPS WITH THIS AI & ML THANG?

Every journey will be different…

My Journey

Graduated as an Electrical Engineer in 1989 went into engineering and programming until I saw this Internet thing and thought this may be big.

Towards the end of 2014, I had the same thought around AI & ML.

My journey started with a Microsoft Course which focused on the techniques not the math.

Did three math heavy modules:

- Caltech (Introductory ML course)
- Columbia AI
- Columbia Machine Learning

Seconds out!

Etienne Greeff
@etienne_greeff

HOW DO I GET TO GRIPS WITH THIS AI & ML THANG?

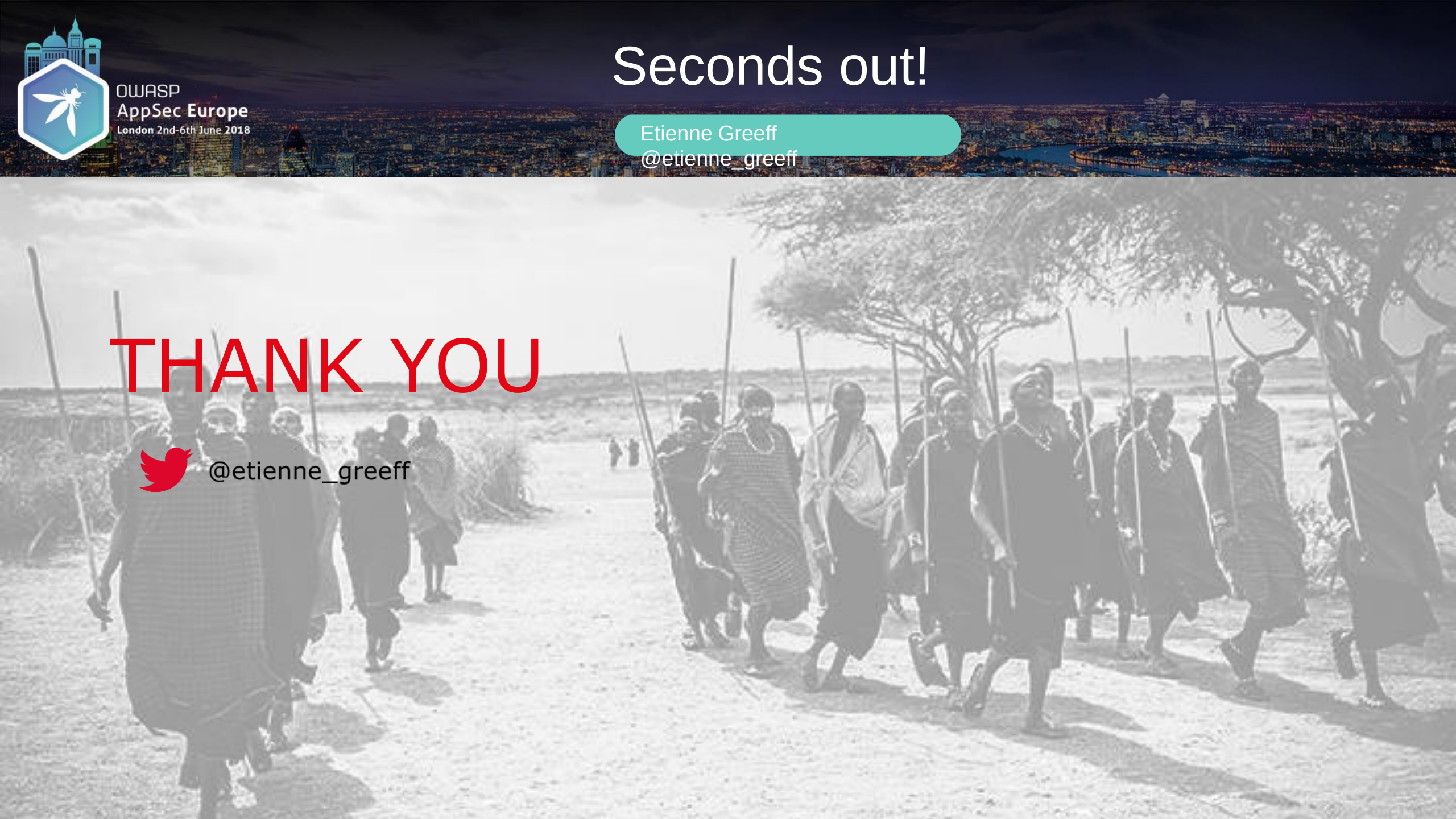
Executive	Development Manager & Strategist	Implementer
Machine Learning for Executives	Brush-up on statistics Brush-up on linear algebra	Do Statistics refresher course. Do Linear Algebra refresher course. Make sure you can truly make python/R sing
Hire the right people. NB. You need both subject matter expert and AI/ML person	Microsoft introductory course	Columbia or MITx courses
Work with companies that solve specific problems. Be cautious of whole solutions…	Specific Microsoft courses depending on requirement	Follow up with specialisation course i.e. deeplearning.ai
	For extra credit Coursera course from Andrew NG	Find a problem and solve it

Seconds out!

Etienne Greeff
@etienne_greeff

TAKEAWAYS

- Understand the new threat models that AI & ML may introduce
- Educate yourself on the subject. This will become as core to most jobs as computing is today
- Ask the right questions to your suppliers
  - What problem does your software solve
  - How does it solve it?
  - What is the false positive rate if anomaly detection and how many alerts can I expect
- Be very sceptical to people making bold claims:
  - World class academics (How many papers have they published?)
  - Protects against cyberattack
  - Protects against attacks you haven’t seen before
  - Fully automates your defence
- Focus on problems that you have a clear problem statement for
- Either build a team or partner with somebody with a track record
- Keep track of the latest developments on arxiv.org or get somebody in your team to do so


OWASP
AppSec Europe
London 2nd-6th June 2018
Seconds out!
Etienne Greeff
@etienne_greeff
THANK YOU
@etienne_greeff