Seconds out!

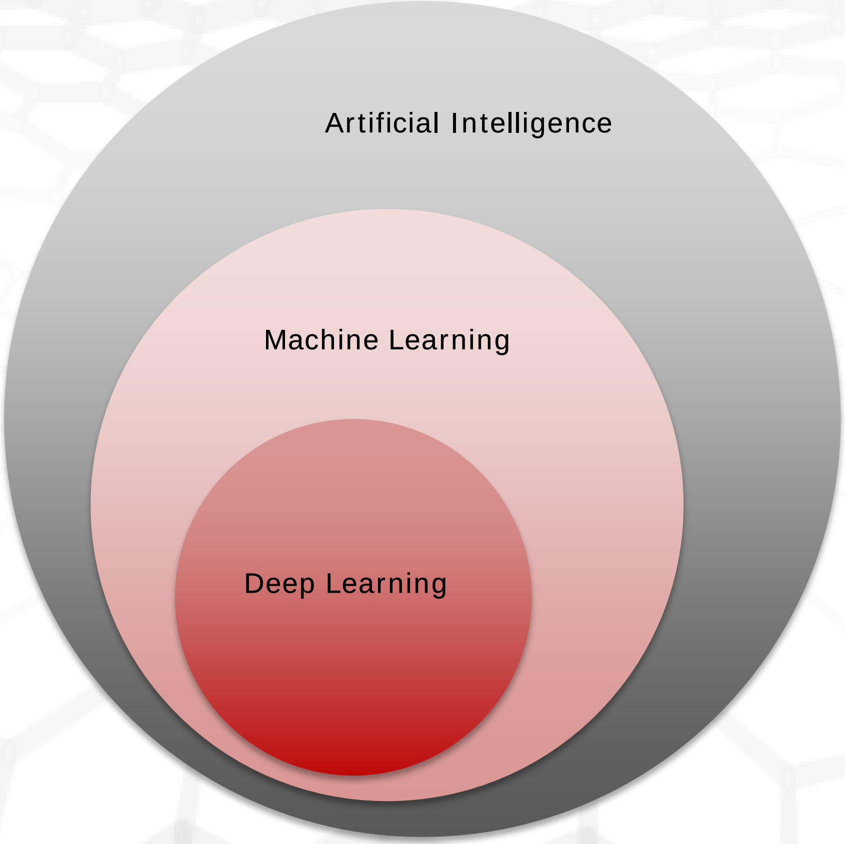
When algorithms don’t play nice with our applications and lives

with Etienne Greeff

@etienne_greeff

AI & ML ARE THE SAME YET DIFFERENT ?

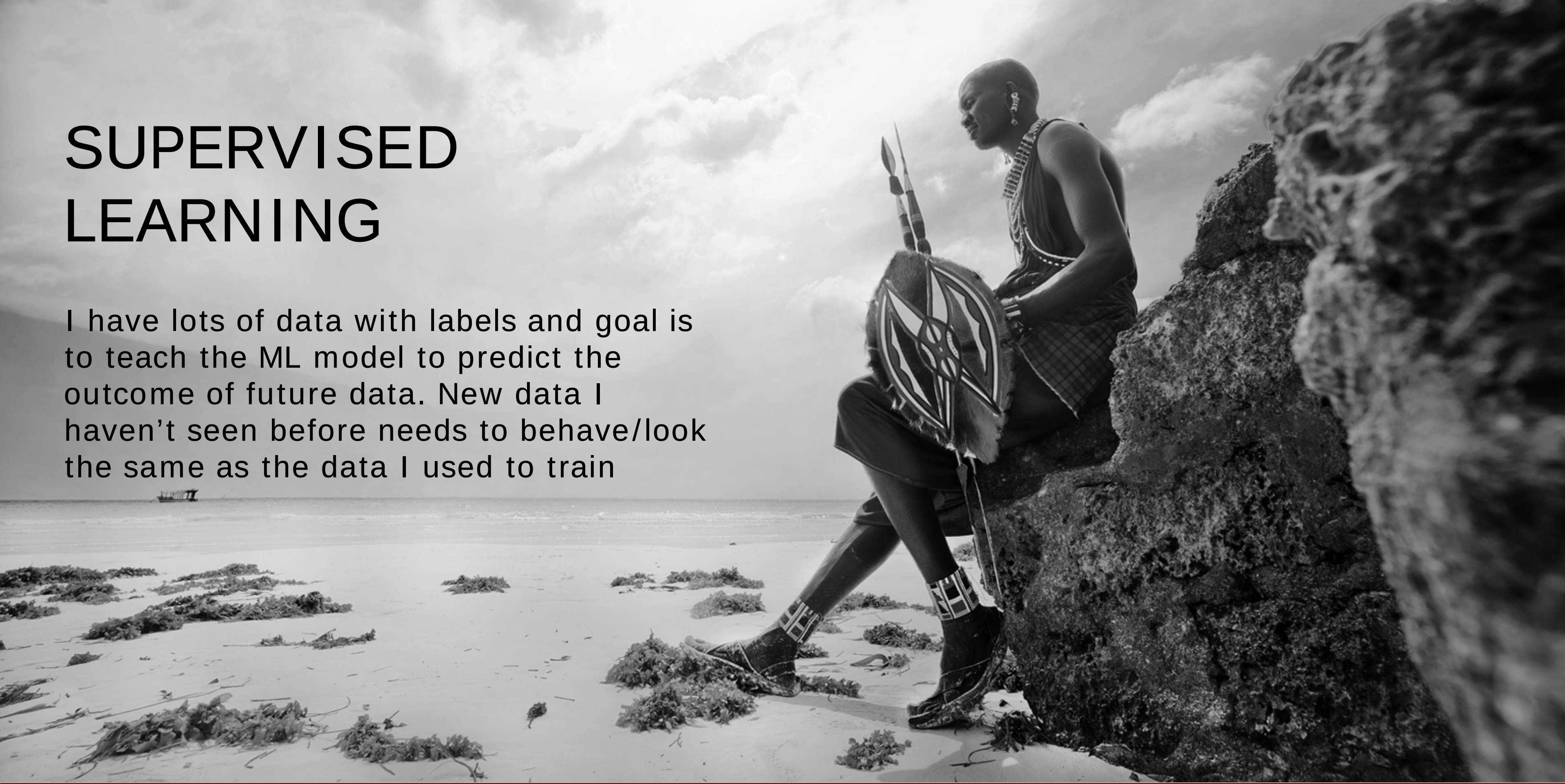
AI seems to be the encompassing marketing term



TWO BROAD TYPES OF ML ALGORITHMS:

Supervised
Have a lot of data and train a mathematical model to predict outcomes for example classify traffic as suspicious or non suspicious

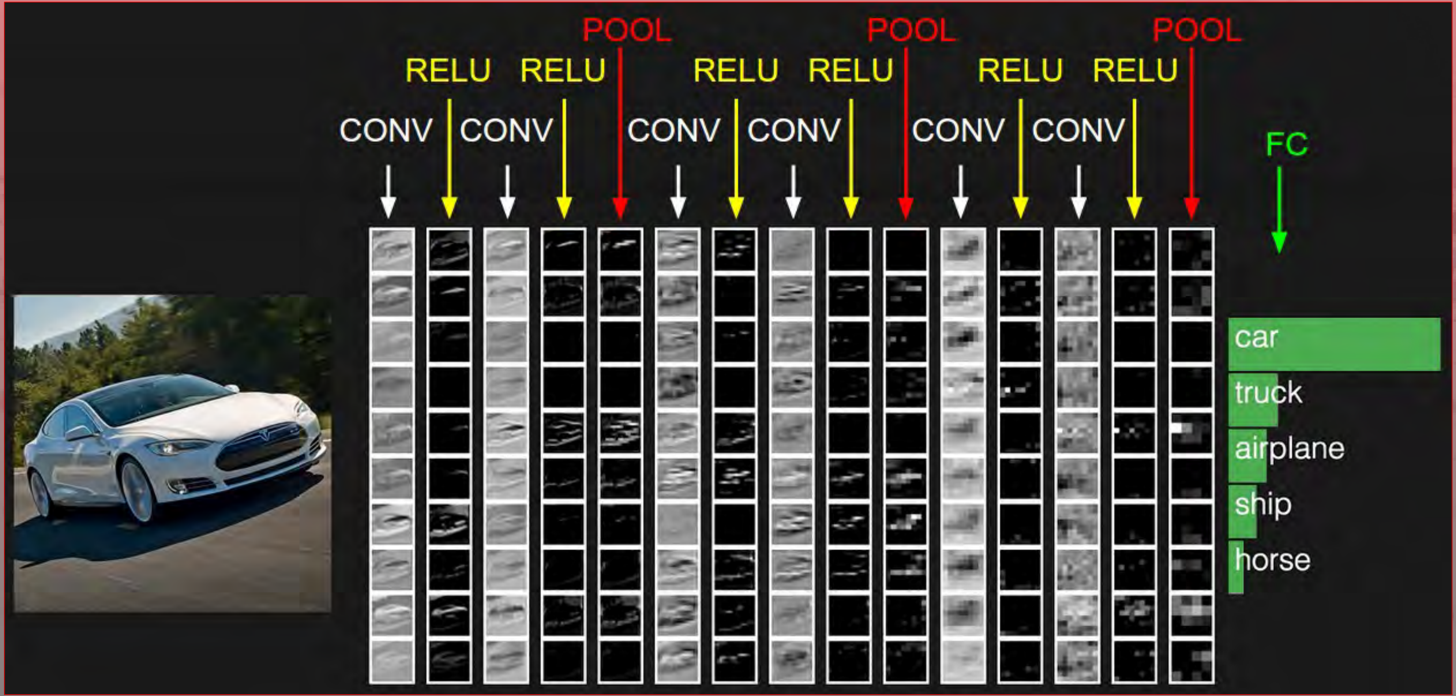
Unsupervised
Don't have labels for my data and am trying to detect structure in my data for example which users are behaving the same way when accessing my application, and more importantly which users behave different from all the others.



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

EXAMPLE OF SUPERVISED LEARNING : NEURAL NETWORK

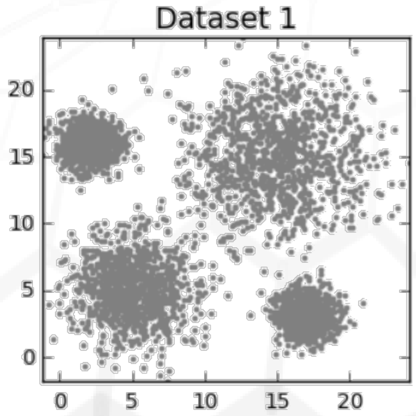


UNSUPERVISED LEARNING

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 suspicious

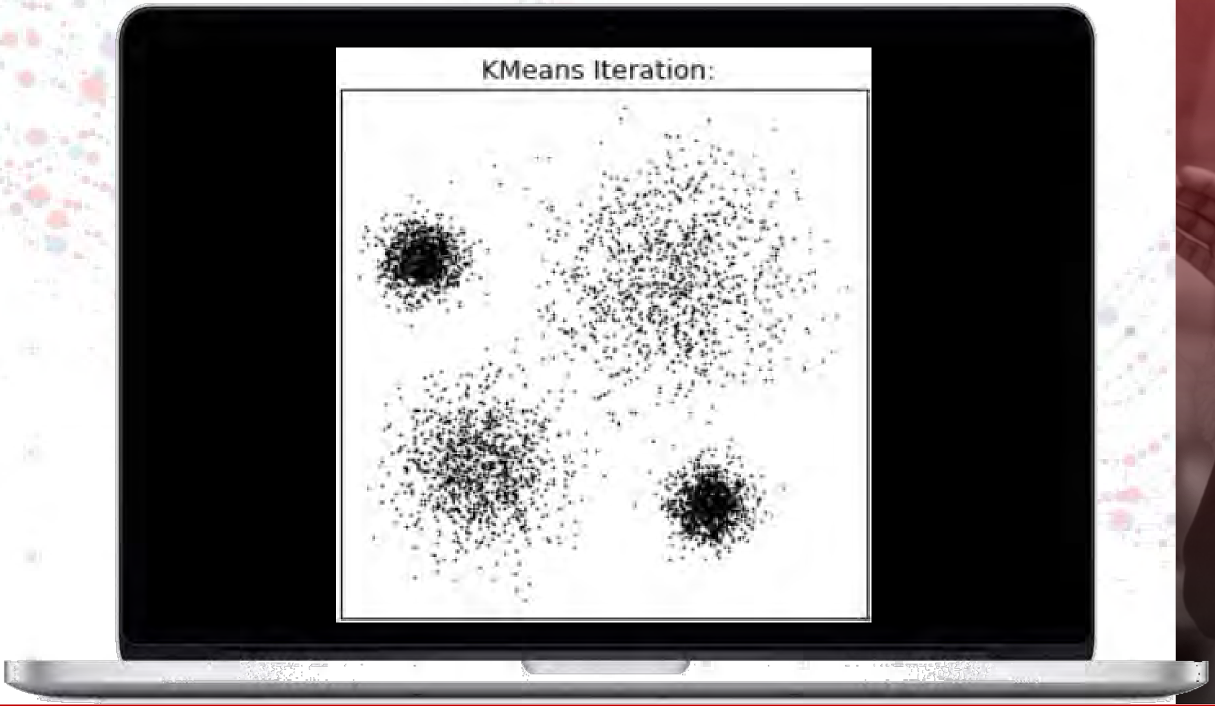
UNSUPERVISED LEARNING GROUPING SIMILAR POINTS TOGETHER USING TWO APPROACHES

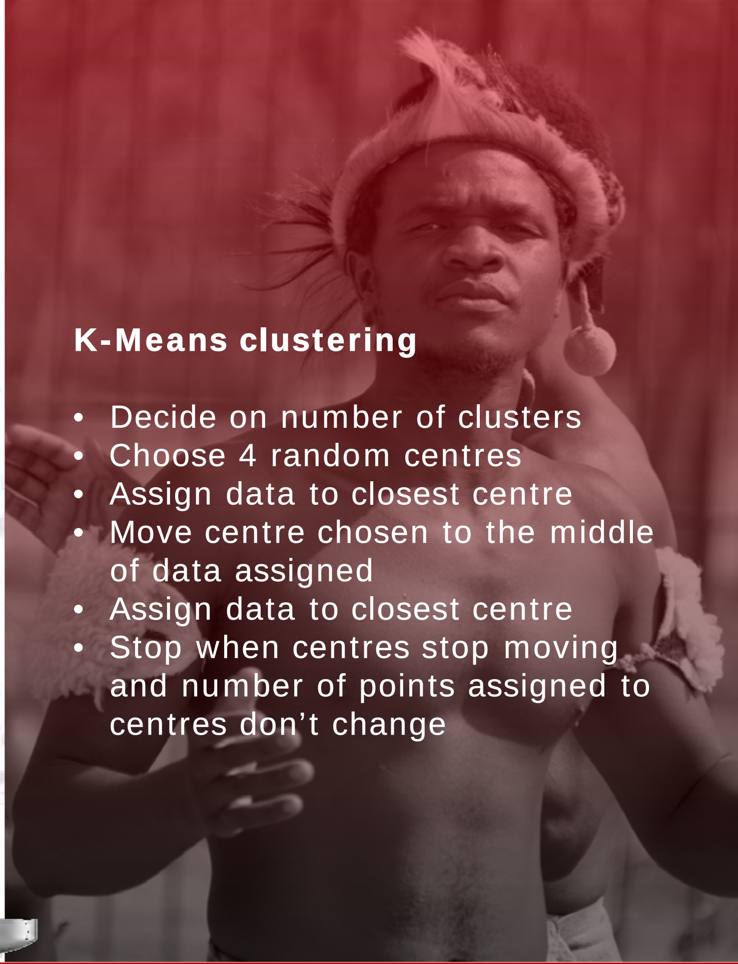
- Discriminative – does not rely on prior knowledge of data
- Generative – assumes the data was generated using a known mathematical function (prior)



UNSUPERVISED LEARNING

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





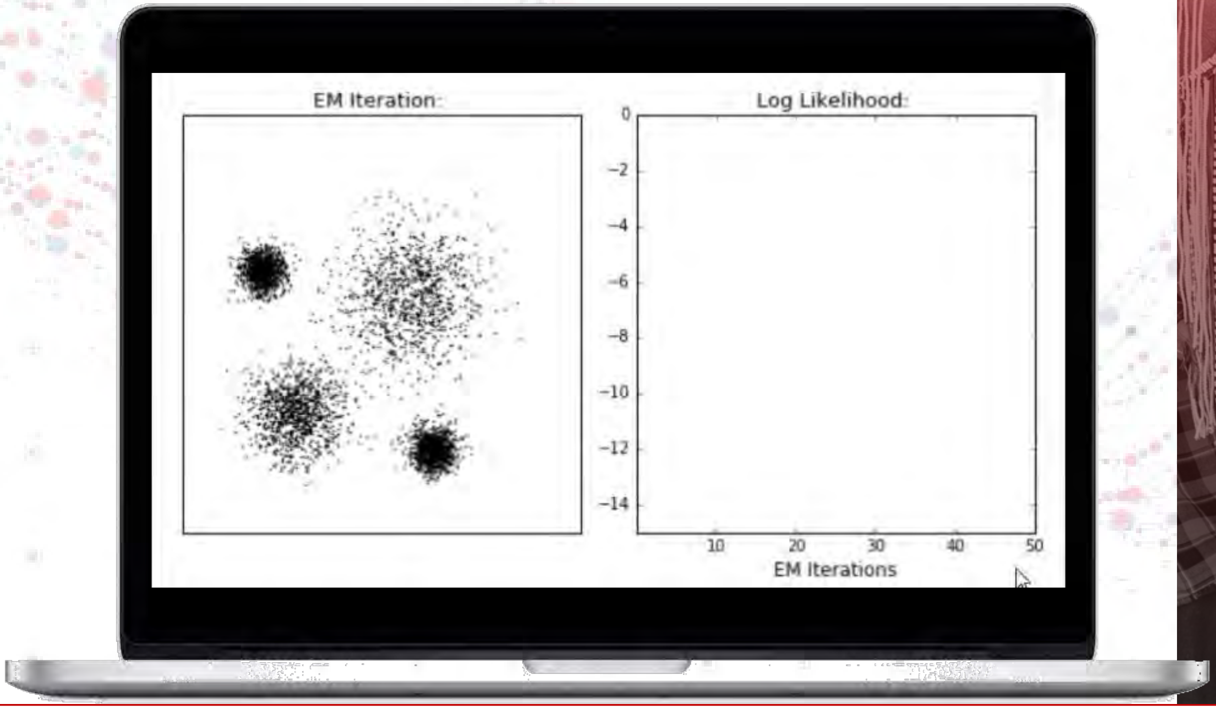
K-Means clustering

- 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

WHAT IS A GAUSSIAN?

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

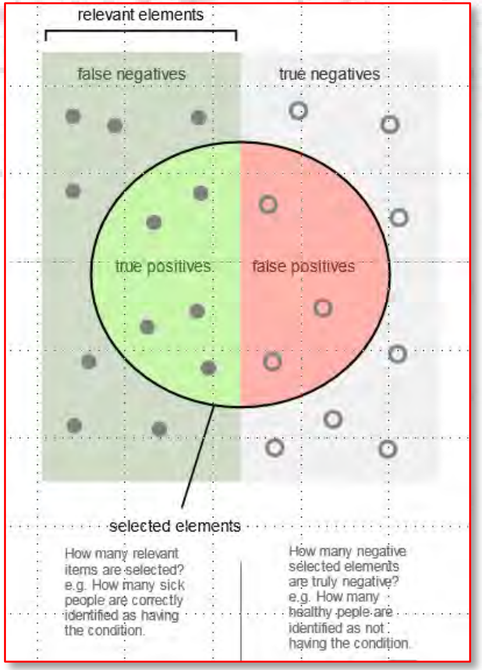
SO WHO IS THIS BAYES GUY THAT IS GOING TO SOLVE ALL OUR PROBLEMS?

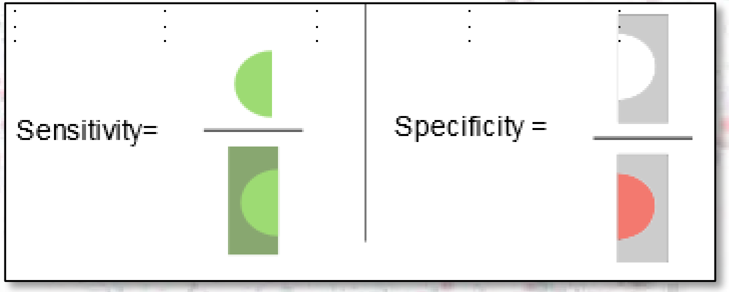
Thomas Bayes (1702-1761)

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

My new belief ∝ What I am actually seeing x What I believed before

SO, HOW GOOD IS IT?





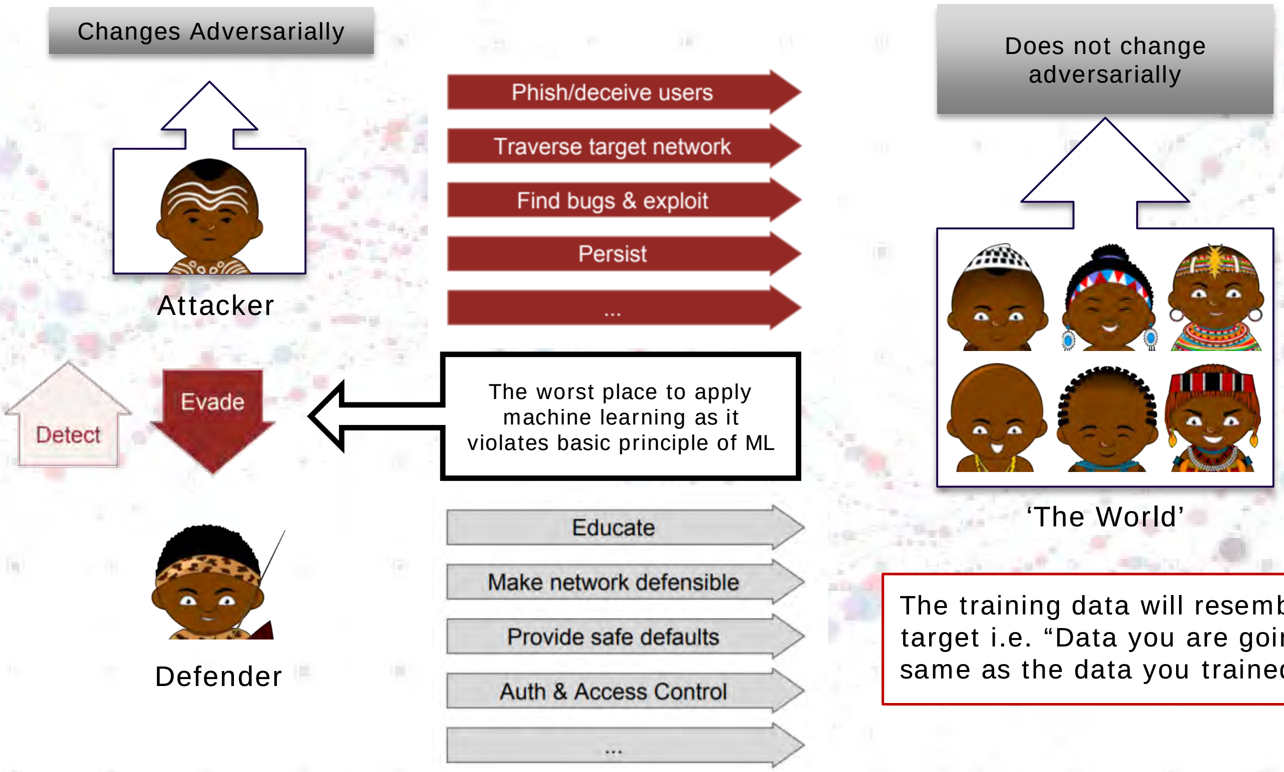
Sensitivity: (also called the **true positive rate**) measures the proportion of actual positives that are correctly identified as such (e.g., the percentage of anomalous log entries which are correctly identified as anomalous).

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).

SO…WILL AI & ML PROTECT US?

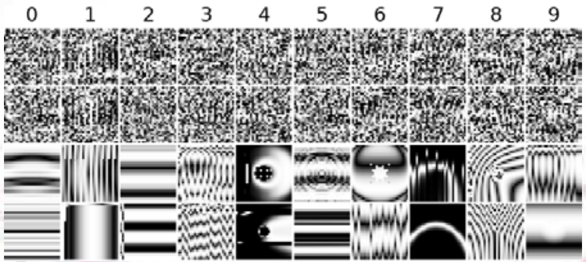
Most current solutions are deployed in the wrong place in the wrong way!

SECUREDATA
TRUSTED CYBERSECURITY EXPERTS

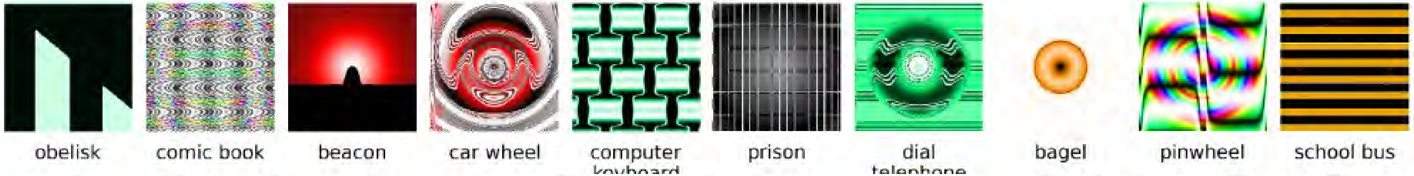


The training data will resemble their real-world deployment target i.e. "Data you are going to work on needs to look the same as the data you trained your ML algorithm on."

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.

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

AI & ML AS A TOOL

- Machine Learning is a tool, not a solution
- People who tell you it's a solution are peddling snake oil
- Organisations doing pioneering work are those creating 'tools' to solve specific problems:
    - Microsoft
    - Google
    - Amazon
    - Symantec
    - Cylance
    - Specialist firms like SecureData, Endgame Systems and others

AI & ML AS A TOOL

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

Here are two examples:

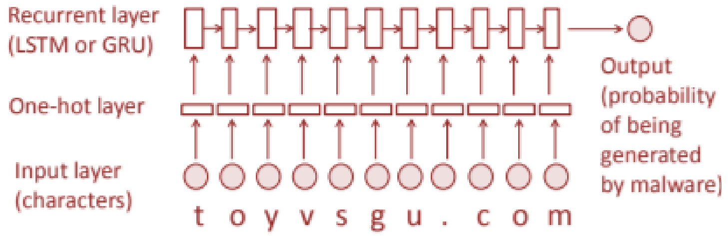
- Identify calls to C&C servers from customer networks
- Analysing large volumes of log messages

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.

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



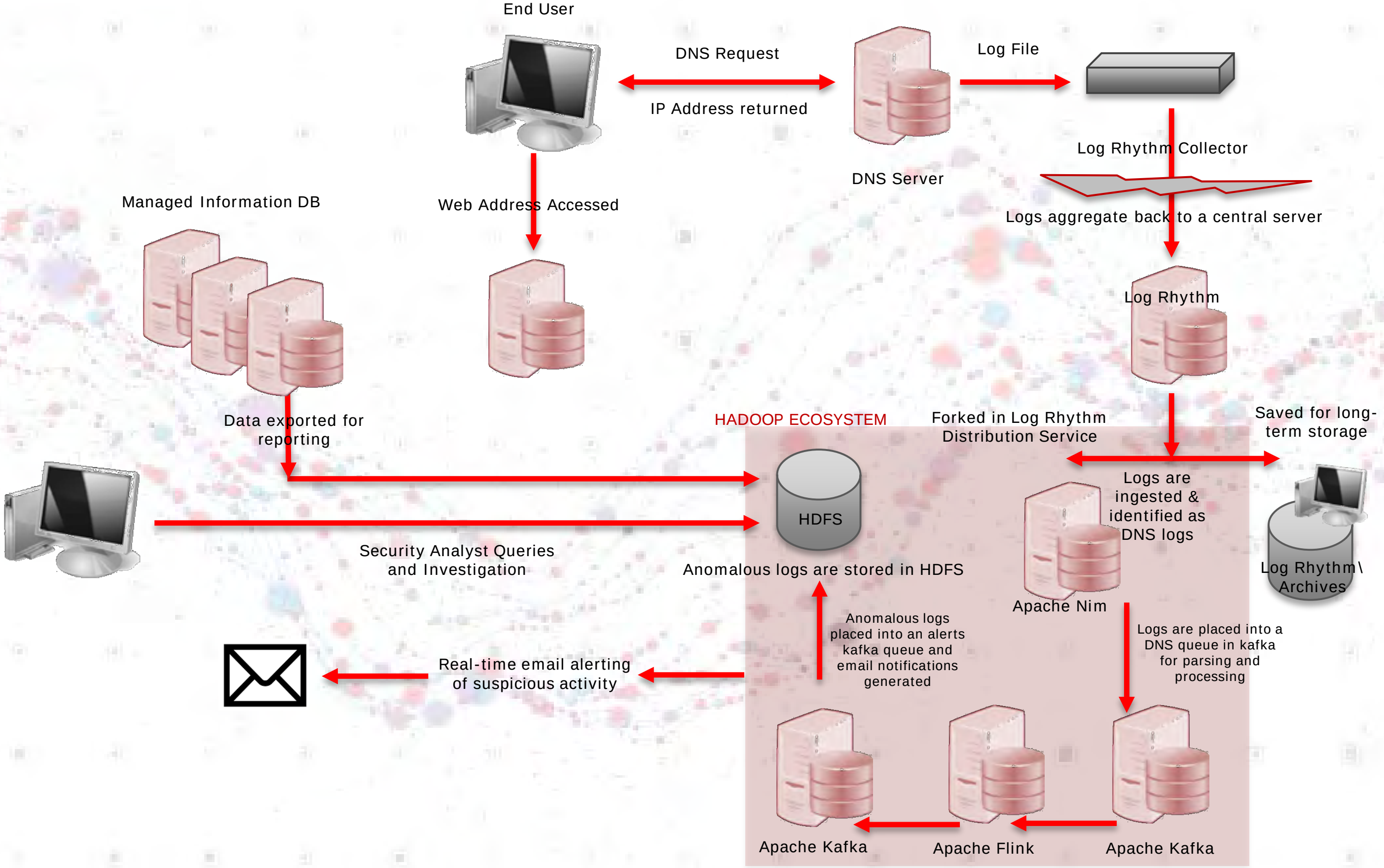

```
model = Sequential()
model.add(Embedding(max_features, 128, input_length=maxlen))
model.add(LSTM(128))
model.add(Dropout(0.5))
model.add(Dense(1))
model.add(Activation('sigmoid'))

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

DETECTING C&C PRODUCTION SETUP



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	0.99910694360733303

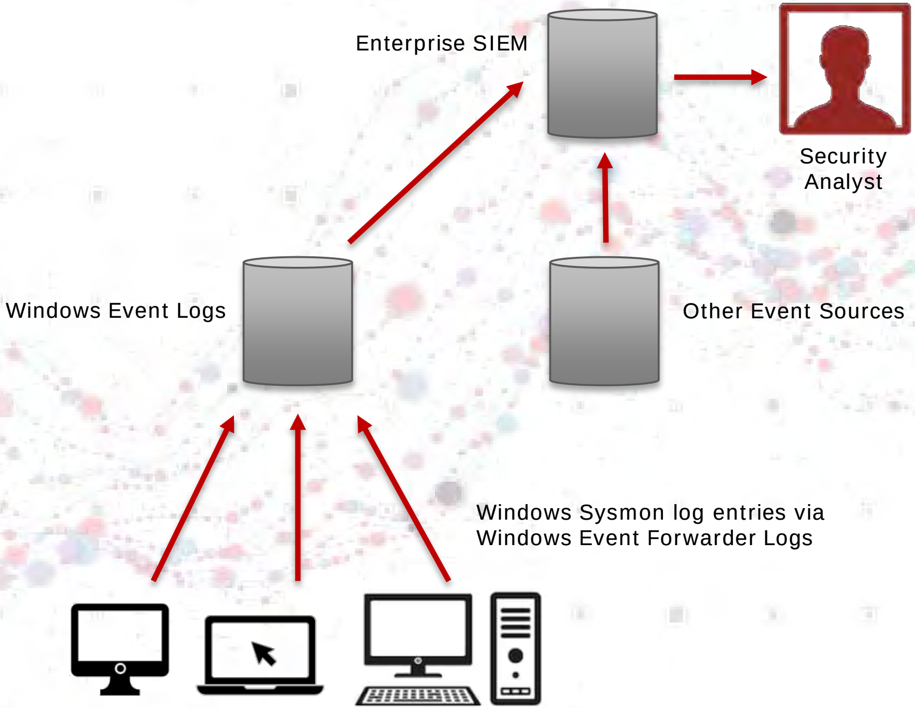
EXAMPLE 2: ANALYSING LARGE VOLUMES OF DATA

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

DEALING WITH 25,000 LOG ENTRIES PER DAY

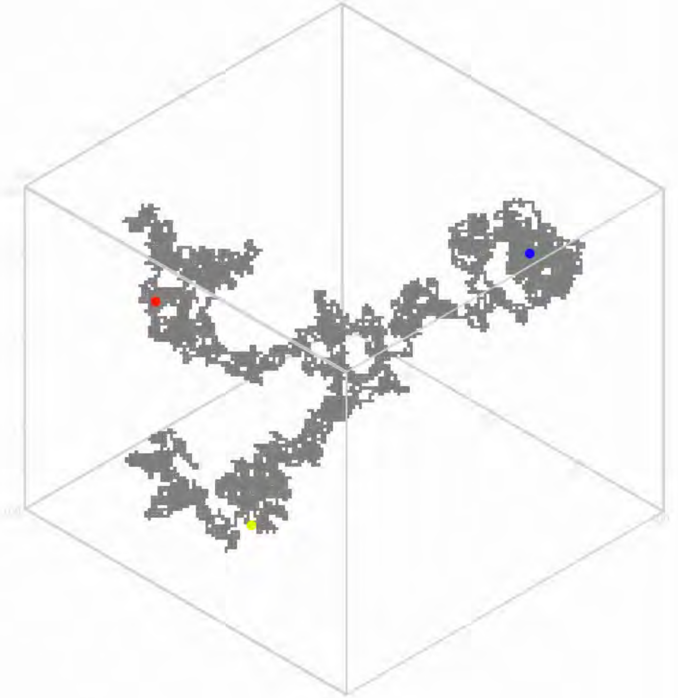
- We use sysmon monitoring on endpoints to log pertinent events
- Typically will receive about 25,000 entries per customer, per day
- It's not possible to go through all the entries
- Would like to group similar entries together so we can analyse quickly
- Would be good if the system can get smarter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



THE RULES

Label ID	Label	Comment	F
0	Looks like GotoMeeting	Looks like G2MLauncher	L
1	Looks like Nslookup	Network Recon Tool from the Command line I suspicious	
2	Looks like Netstat	Network Recon Tool from the Command line I suspicious	
3	Looks like ipconfig	Cmder is a known terminal emulator for Windows	
4	Looks like net.exe	Looks like someone trying to start, stop, pause or restart a service	
5	Looks like common Splunk Powershell	Looks like a standard Splunk Powershell App	
6	Benign DCOM Server Process Launcher service	The DCOMLAUNCH service launches COM and DCOM servers in response to object activation requests.: C:\Windows\system32\svchost.exe -k DcomLaunch	
7	Weird Powershell SVCHost thing.	Not sure what's happening here - we need to check it out.	
8	Benign Symantec Endpoint Stuff - SYS File	c:\program files (x86)\symantec\symantec endpoint protection\14.0.3897.1101.105\bin\ccsvchst.exe	
9	Benign office accessing an inf file	Office opening a file from \appdata\local\temp\	
10	Looks like Word opening a Word document	We would expec to see Office running from Startup	
11	Known app from Expected Location	We would expec to see Office running from Startup	
12	Bening looks like Outlook spawning Chrome	Probably benign Outlook spawning Chrome	
13	Looks like Excel spawning Chrome	Looks like Excel Spawning Chrome	
14	Looks like Install Flash Player	Looks like Install Flash Player	
15	Looks like Chrome downloading something - ZONEID	Probably just Chrome downloading something	
16	Bening looks like Outlook spawning Firefox	Probably benign Outlook spawning Firefox	
17	Looks like Firefox downloading something	Probably just Firefox downloading something	
18	Bening looks like Outlook spawning iExplore	Probably benign Outlook spawning iExplore	
19	Benign Opera Browser Update	Benign Opera Browser Update	
20	Looks like Vivaldi downloading something	Vivaldi is a freeware, cross-platform web browser developed by Vivaldi Technologies.	
21	Looks like that weird Powershell thing.	THIS LABEL NEEDS WORK! Powershell. Temp location. Random name. We see this a lot.	L
22	Looks like Windows Scripted Diagnostics	sdiagnhost.exe is run as a standard windows process with the logged in user's account privileges. C:\Windows\System32\sdiagnhost.exe is where this software will be found on a computer. Scripted diagnostics can execute diagnostic packages that are signed by untruste	
23	Looks like Windows Taskhost Powershell	taskhostw.exe file is a software component of Windows service start manager by Microsoft. Taskhostw.exe is part of the Windows 10 operating system, and starts DLL-based Windows services when the computer boots up. Thus, Windows 10 uses the taskhostw.exe file as	
24	Looks like that weird Powershell startupprofiledata-	Looks like that weird startupprofiledata-noninteractive thing	
25	Looks like Slack accessing the downloads folder	Looks like Slack accessing the downloads folder	
26	Looks like Teams accessing the Downloads folder	Looks like Teams accessing the Downloads folder	
27	Looks like a Flash Macromedia Installer	Looks like a Flash Macromedia Installer	
28	Looks like Firefix accessing the Cache	Looks like Firefix accessing the Cache	
29	Looks like Chrome writing tempfiles	Looks like Chrome writing tempfiles to the downloads folder	
30	Looks like Chrome sofware reported tool	Looks like Chrome sofware reported tool	
31	Benign Semantec endpoint accessing an inf file	Office opening a file from \appdata\local\temp\	
32	Benign looks like MS PickerHost.	The PickerHost.exe is a File Picker UI Host. This file is part of Microsoft Windows Operating System. PickerHost.exe is developed by Microsoft Corporation. It's a system and hidden file. PickerHost.exe is usually located in the %SYSTEM% folder and its usual size is 25,680 b	
33	Looks like a OneNote Link File	Looks like a OneNote Link File	
34	Looks like Google updater	Looks like Google updater	
35	Looks like Citrix Service Updater - BAT file	Looks like Citrix Service Updater	
36	Excel opening an Excel file from downloads - XLSX fil	Excel opening an Excel file from downloads.	
37	Looks like ESIF	Looks like ESIF	
• • •			
94	Looks like OneNote talking out	Looks like OneNote talking out	
95	Looks like Word accessing a DOCX file	Looks like Word accessing a DOCX file	
96	Looks like MS Browser Broker : ZONEIDENTIFIER file	browser_broker.exe application is part of Windows/System32, associated with dealing with downloads.	
97	Excel opening an Excel file from downloads - TMP file	Excel opening an Excel file from downloads.	
98	Looks like Outlook downloading a file - XLSX	Outlooking opening an XLSX file	
99	Benign Symantec Endpoint Stuff - ZIP File	c:\program files (x86)\symantec\symantec endpoint protection\14.0.3897.1101.105\bin\ccsvchst.exe	
100	Excel opening an XML file	Excel opening an XML file	
101	Looks like Powershell accessing a DAT file	THIS LABEL NEEDS WORK! Powershell. Temp location. Random name. We see this a lot.	
102	Looks like Windows Diagnostics Troubleshooting Wi	he original msdt.exe from Microsoft is an important part of Windows, but often causes problems. Msdt.exe is located in the C:\Windows\System32 folder or sometimes in a subfolder of C:\Windows.	

USE MARKOV CHAIN BASED RANDOM WALK SEMI-SUPERVISED CLASSIFIER



THE RESULTS

Probabi	Source	Rule Na	Label ID	Label	Comma	Login	Process	Object	Comma	Process	Object I	Comma	Comma	Login In	SHost
1	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	sophiehalt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	stuartblt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	natalieclt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	garethbelt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	davidhalt.mis-cds.local
0.98864															t.mis-cds.local
0.98864															t.mis-cds.local
0.98864															blt.mis-cds.local
0.98864															egrlt.mis-cds.local
0.98864															t.mis-cds.local
0.98864															t.mis-cds.local
0.98864															hrlt.mis-cds.local
0.98864															hrlt.mis-cds.local
0.98864															hrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	hannahrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	neilblt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	samhlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	samhlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	etiennegrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	etiennegrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local
0.9886439	<aie v="1"><_0 D	AIE: SD: Sy	45	Weird. Looks like Excel connecting out.	Looks like Excel conne		NoProcess	c:\program	NoComma	NOPROCES	EXCEL.EXE	NoComma	NoComma	<Blank>	andyrlt.mis-cds.local


```
<aie v="1"><_0 DHost="1|0|9.9.9.9|-1|dns.quad9.net" NormalMsgDate="2018-07-03 11:49:21"
NormalMsgDateLower="2018-07-03 11:49:21" NormalMsgDateUpper="2018-07-03 11:49:22"
Object="c:\program files\microsoft office\root\office16\excel.exe" RootEntityID="8"
RuleBlockType="1" SHost="1|0|10.10.23.207|-1|etiennegrlt.mis-cds.local" />
<_ AIERuleID="1000000231" DateEdited="2018-03-29 08:35:51" /></aie>
```


Identifies log entries that are similar to other rules but without explicit rules

Having to analyse 28 entries manually now rather than trawling through 25,000

PEERING INTO THE FUTURE

SECUREDATA
TRUSTED CYBERSECURITY EXPERTS

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
Fuzzing applications for fun and profit & penetration testing	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
Government involvement/power projection using Cyber	Government involvement/power projection using AI & ML

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.

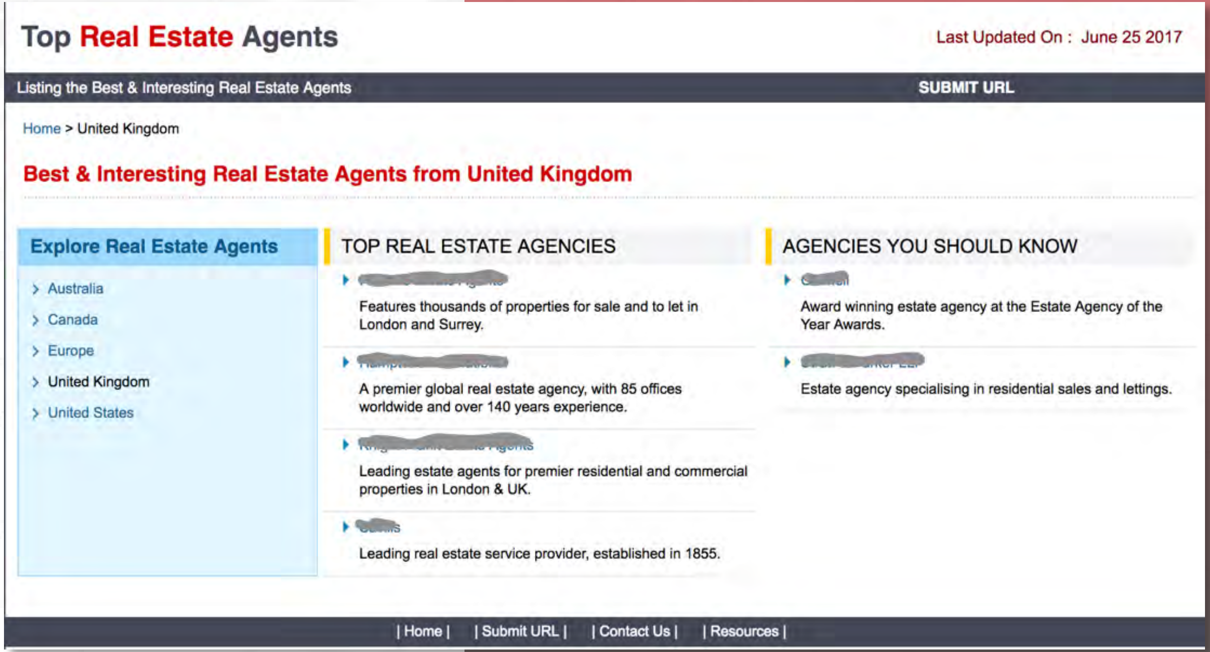
Home buyers stand to lose thousands in new cyberattack

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

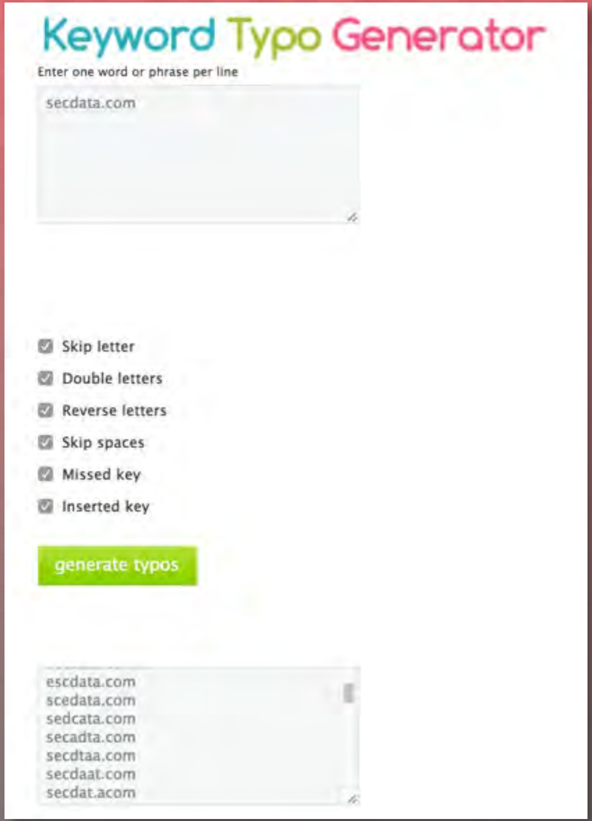
TYPO SQUATTING

2 Enumerate the players



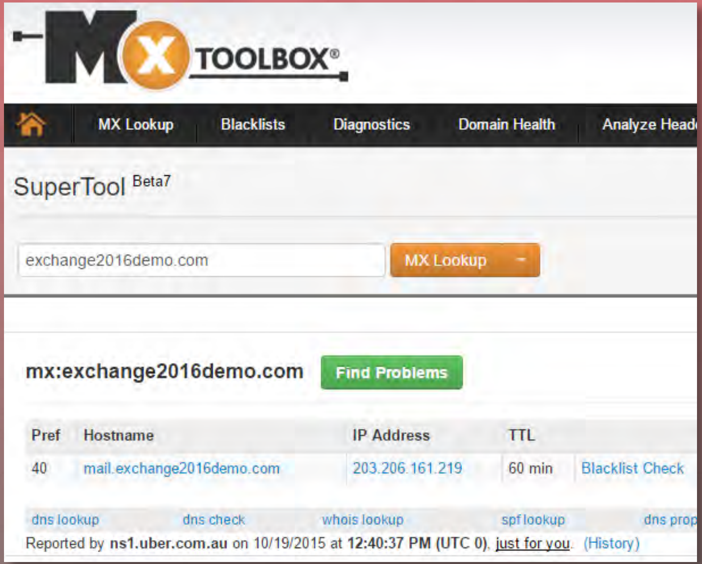
TYPO SQUATTING

3 Generate the Typos



TYPO SQUATTING

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



SECUREDATA
TRUSTED CYBERSECURITY EXPERTS

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

AI EQUIVALENT OF TYPO SQUATTING

A.k.a. Voice Squatting and Voice Masquerading

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

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

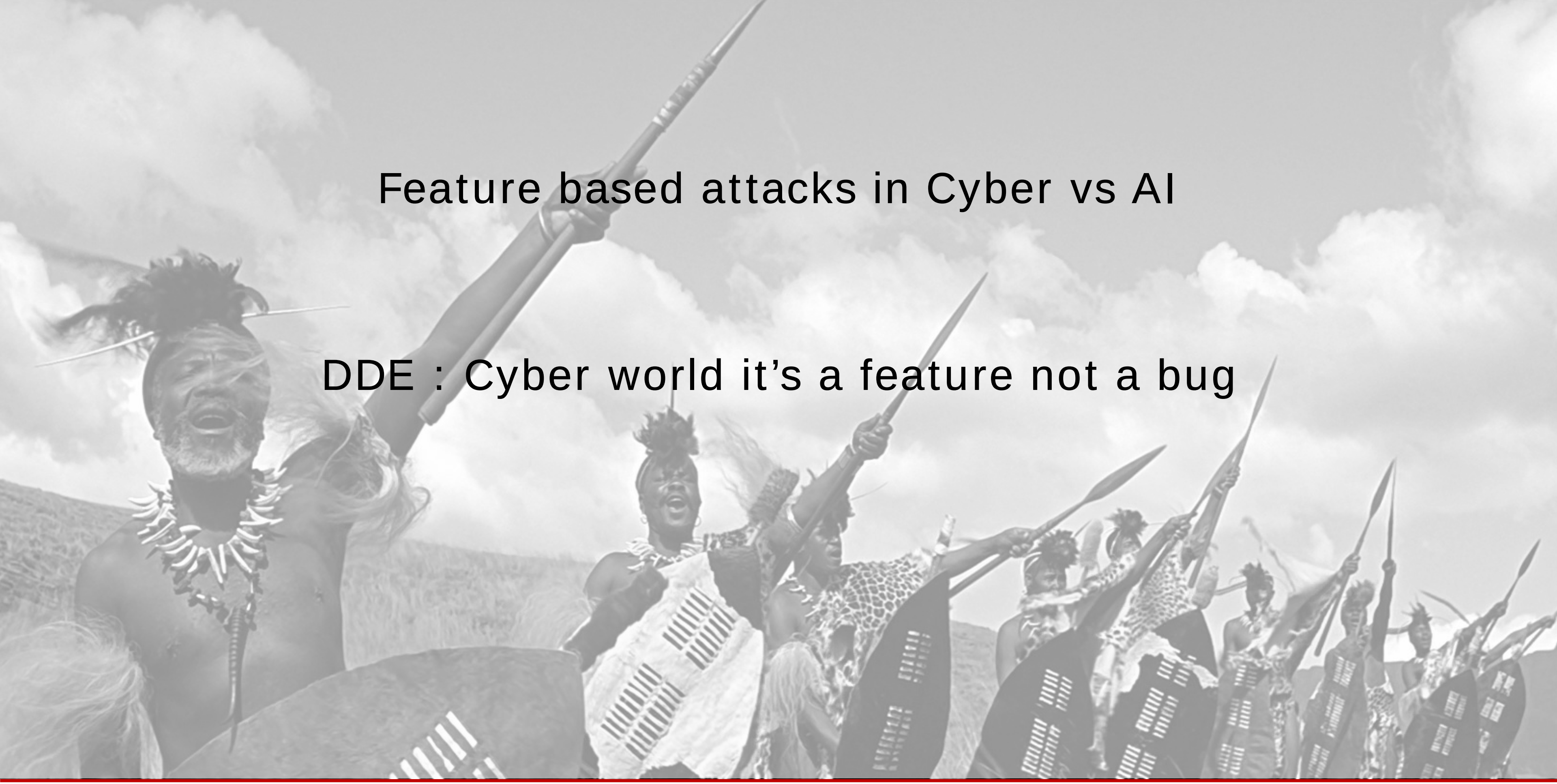
Voice Squatting:

- 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

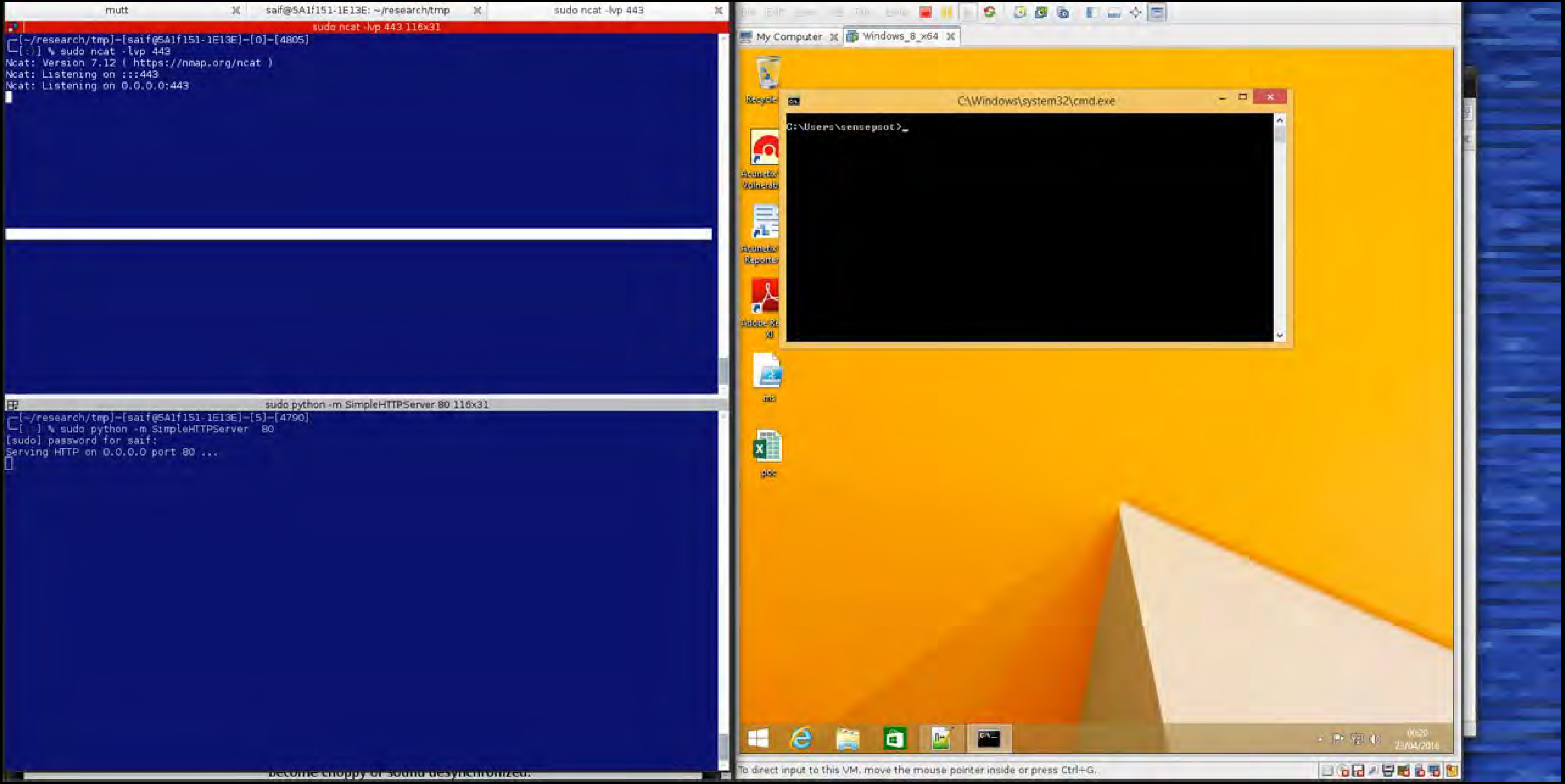
- When the trust system on the VPA is abused
- The VPA relies on the current running skill (which may be malicious) to stop
- The malicious skill is then in a position to gather all types of confidential information as it continues running
- Real life tests show this is possible and people don't pay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

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



Feature based attacks in Cyber vs AI

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

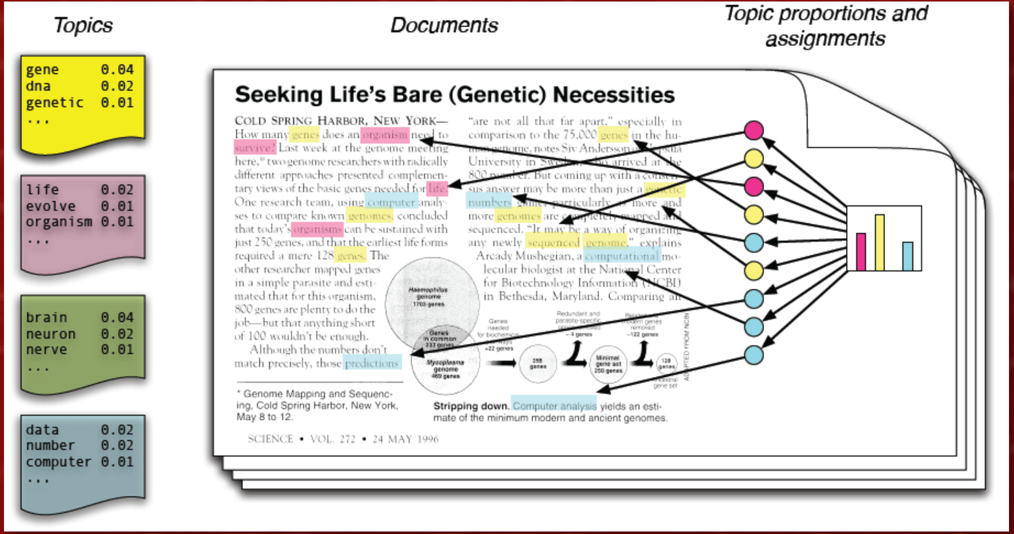

mutt
saif@5A1f151-1E13E: ~/research/tmp
sudo ncat -lvp 443
sudo ncat -lvp 443 116x31
[~/research/tmp]-[saif@5A1f151-1E13E]-[0]-[4805]
[:)] % sudo ncat -lvp 443
Ncat: Version 7.12 (https://nmap.org/ncat)
Ncat: Listening on :::443
Ncat: Listening on 0.0.0.0:443
sudo python -m SimpleHTTPServer 80 116x31
[~/research/tmp]-[saif@5A1f151-1E13E]-[5]-[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\sensepost>_
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.

Feature based attacks in Cyber vs AI

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

TOPIC MODELLING FOR FUN AND PROFIT ON DESKTOPS

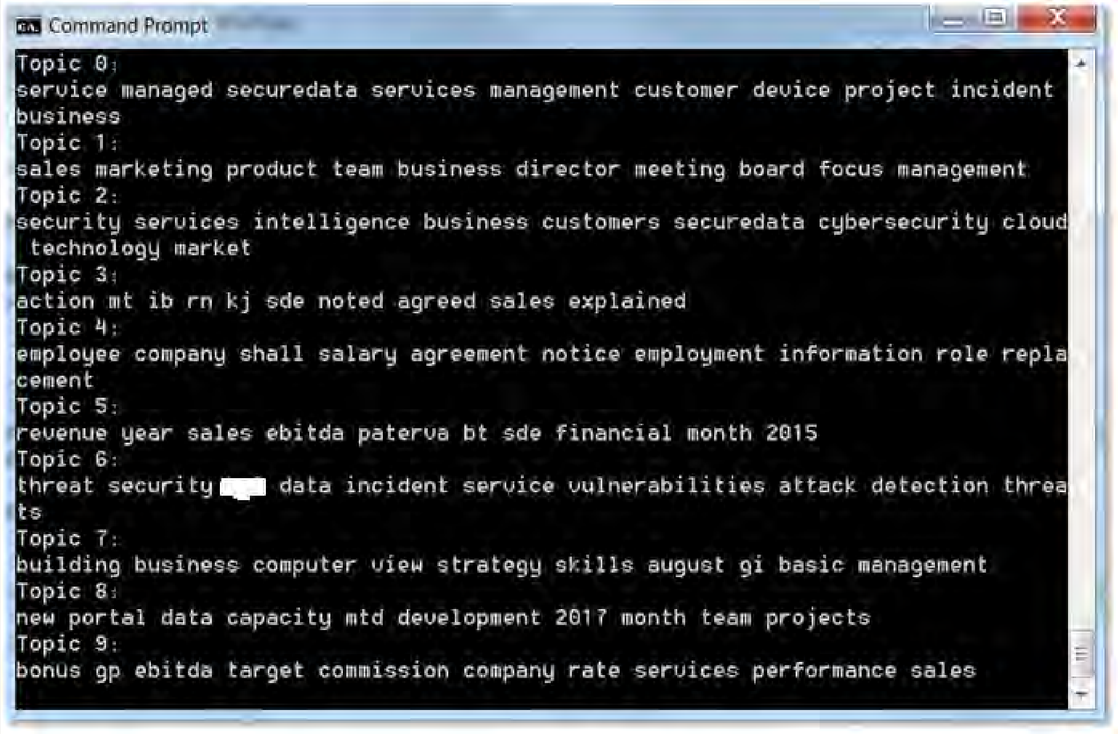
Topic modelling is the process of analysing which words are used together the most in the most common ways in other words what topics does this bunch of documents discuss using which words.



SECUREDATA
TRUSTED CYBERSECURITY EXPERTS

TOPIC MODELLING FOR FUN AND PROFIT ON DESKTOPS

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



KEY 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 of 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 automate 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

THANK YOU

QUESTIONS?

@etienne_greeff