

# Challenges of the IoT and Big Data

Bernard Butler

Emerging Networks Laboratory, TSSG, WIT



Waterford Institute of Technology  
INSTITIÚID TEICNEOLAÍOCHTA PHORT LAIRGE

# Overview



1. Introduction  
Some definitions
2. Exercise
3. The Context and History
4. Expertise and Roles
5. Is the IoT and Big Data always a force for good?
6. Conclusions

## Selected definitions

**IoT:** A network of pervasive connected objects able to collect and exchange data from embedded sensors, with the infrastructure and services to support them. (Various)

**Big Data:** High volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. (Gartner 2012)

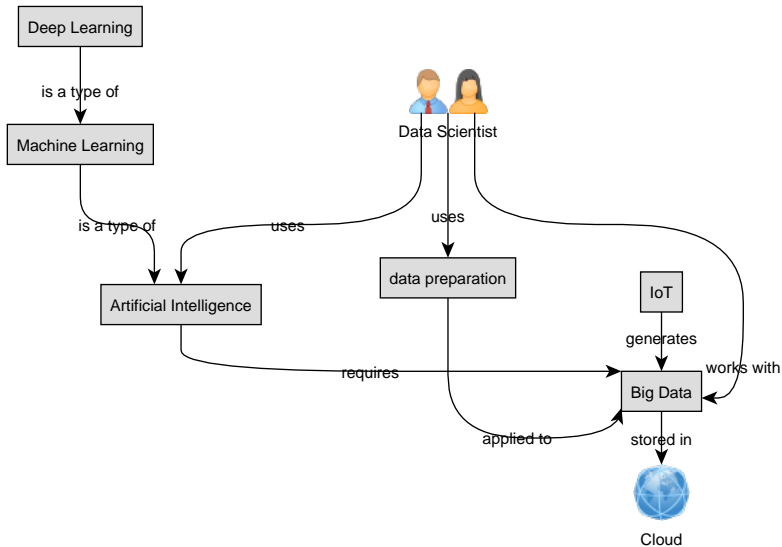
**Data Scientist:** can ask the right questions, {generate} and consume the results of analysis of Big Data effectively. (McKinsey 2011)

**Artificial Intelligence:** the capability of a machine to imitate intelligent human behavior (Webster 2017)

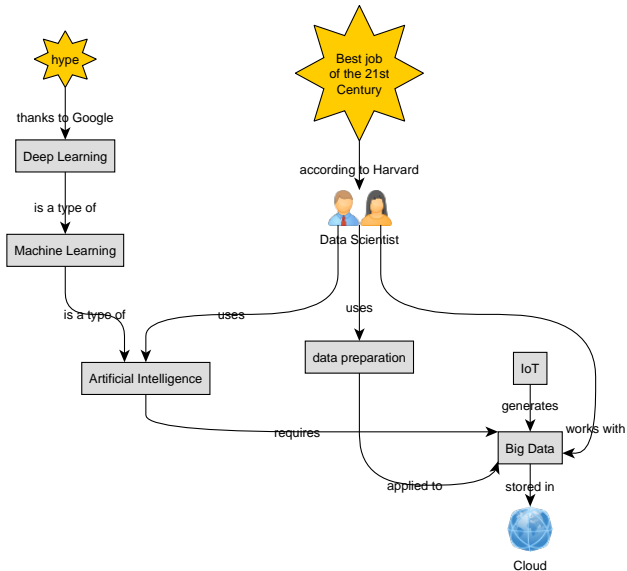
**Machine Learning:** Branch of computer science {and related fields} that gives computers the ability to learn without being explicitly programmed. (Samuel 1959)

**Deep Learning:** Use of very large neural networks with many layers of “neurons” that can be trained to generate robust models of their input, whose classification performance scales with the amount of data supplied. (Various)

# Relationships between terms



# Annotated Relationships between terms!



## Interlude: Examples of Big Data



### Exercise

In pairs, please consider (real world) processes generating *Big Data*.  
Can you come up with 3 examples in 2 minutes?

## Prehistory, or more than 10 years ago...

### Data Generation

- Transactions (bank, retail)
- Activity, e.g., texts
- Basic e-commerce

### Data Processing

- Databases, SQL, stored procedures
- Consultants, system integrators
- Proprietary statistical software

### Data Analysis

- Reporting: looking back
- Descriptive statistics
- Simple plots

## The first (batch) wave: 2007-2011

### Data Generation

- As before...
- Web activity: comments, etc.
- 360degree view

### Data Processing

- As before...
- NoSQL
- hadoop ecosystem (batch analytics)

### Data Analysis

- As before...
- Personalisation and recommendation
- Predictive Analytics



## The second (streaming) wave: 2012-2015



### Data Generation

- As before. . .
- Social Media!
- IoT (early adopters)

### Data Processing

- As before. . .
- Apache Spark
- R vs. python

### Data Analysis

- As before. . .
- Data understanding
- Weak AI: assistants, etc.

## The current (machine) wave: 2016-?

### Data Generation

- As before...
- Machine-generated (e.g., fake news)
- IoT (mainstream)

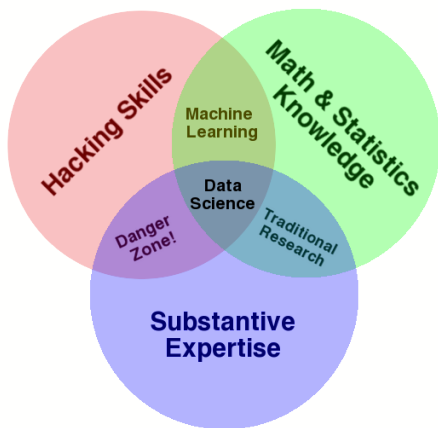
### Data Processing

- As before...
- Microservices: move function to data
- Decoupled databases with schema-on-read

### Data Analysis

- As before...
- Deep learning inflection point
- Visualisation

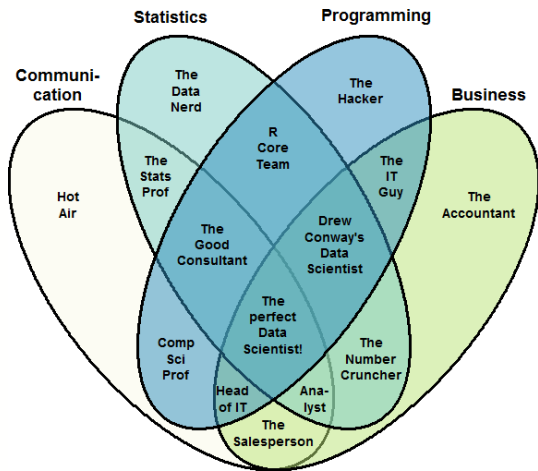
# Drew Conway's 3-set Venn Diagram of Data Science Expertise



*Source:*

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

# Stephan Kolassa's 4-set Venn Diagram of Data Science Expertise



Source: <https://datascience.stackexchange.com/a/2406>

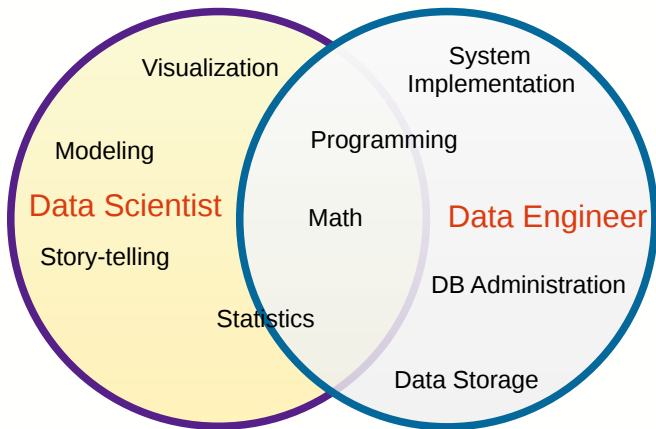
# Gartner suggests the need for a *Citizen Data Scientist*



Source:

<http://www.kdnuggets.com/2016/03/cartoon-citizen-data-scientist.html>

# Data Scientist vs Data Engineer



Source:

<http://101.datascience.community/2014/07/08/data-scientist-vs-data-engineer/>  
Also the traditional roles of *Data Analyst* and *Software Engineer*...

# Complete the following disadvantages of IoT and Big Data



m\_\_\_ s\_\_v\_\_l\_\_\_\_\_

i\_\_\_t\_\_y \_h\_f\_

d\_\_\_c\_ b\_\_n\_\_\_

d\_\_\_\_\_l \_f \_\_r\_\_c\_

b\_\_s

l\_\_\_ o\_ t\_\_\_s\_\_r\_\_\_y

And those disadvantages are...

mass surveillance

identity theft

device botnets

denial of service

bias

lack of transparency





# Conclusions



- Computing is becoming more interdisciplinary
- Research challenges: how to do things better: faster, more accurate, less energy, . . .
- Societal challenges: how to use these new devices, services, interactions, . . .
- Many computing jobs to be filled - so good luck!



## Thank You

Contact:

Bernard Butler,

Research Fellow, TSSG

Waterford Institute of Technology, Ireland

t: +353 (0)51 84 5695

e: [bbutler@tssg.org](mailto:bbutler@tssg.org)



Waterford Institute of Technology  
INSTITIÚID TEICNEOLAÍOCHTA PHORT LÁIRGE