

L'Intelligence Artificielle et la Science des Données – Challenges et Opportunités

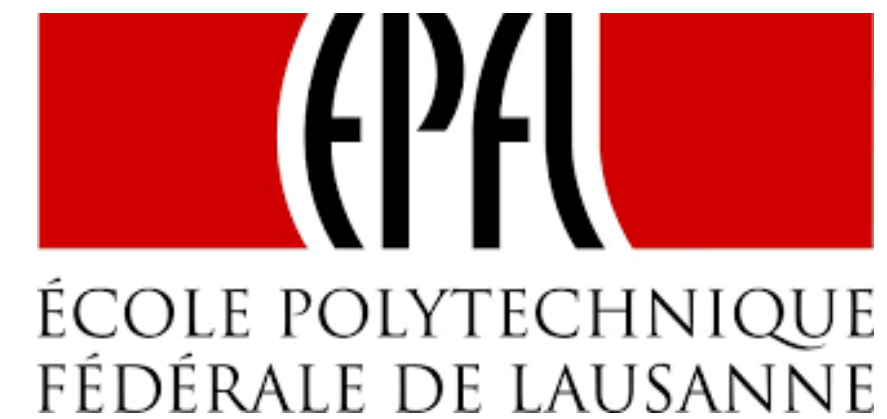
Olivier Verscheure, SDSC @ EPFL & ETH Zurich

About me

Academia



ETH zürich



Lonza

Unil
UNIL | Université de Lausanne
HEC Lausanne

1999

2016



New York



What is Data Science?

Data + Algo

*Big Data /
Data lake*

*Mac
“Dumb” A.I.*



Extract Meaning
and Value out of Data



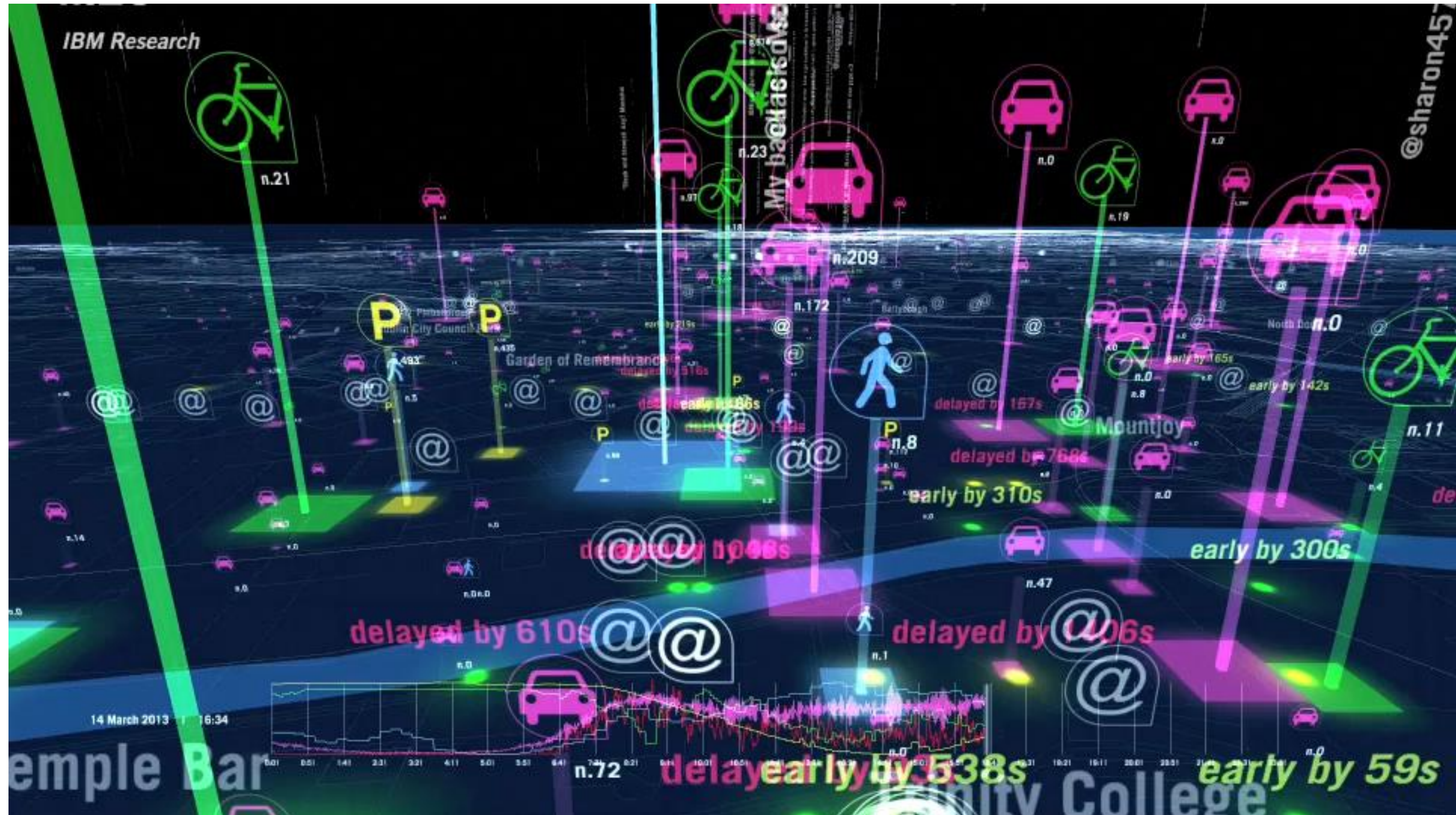
➔ Benefits

What do you see?



O'Connell Bridge / D'ollier St. Dublin City CCTV
8 Apr 2013 18:31:50 GMT Daylight Time

© 2010 Blackwell Publishing Ltd, *Journal of Internal Medicine* 267: 105–112

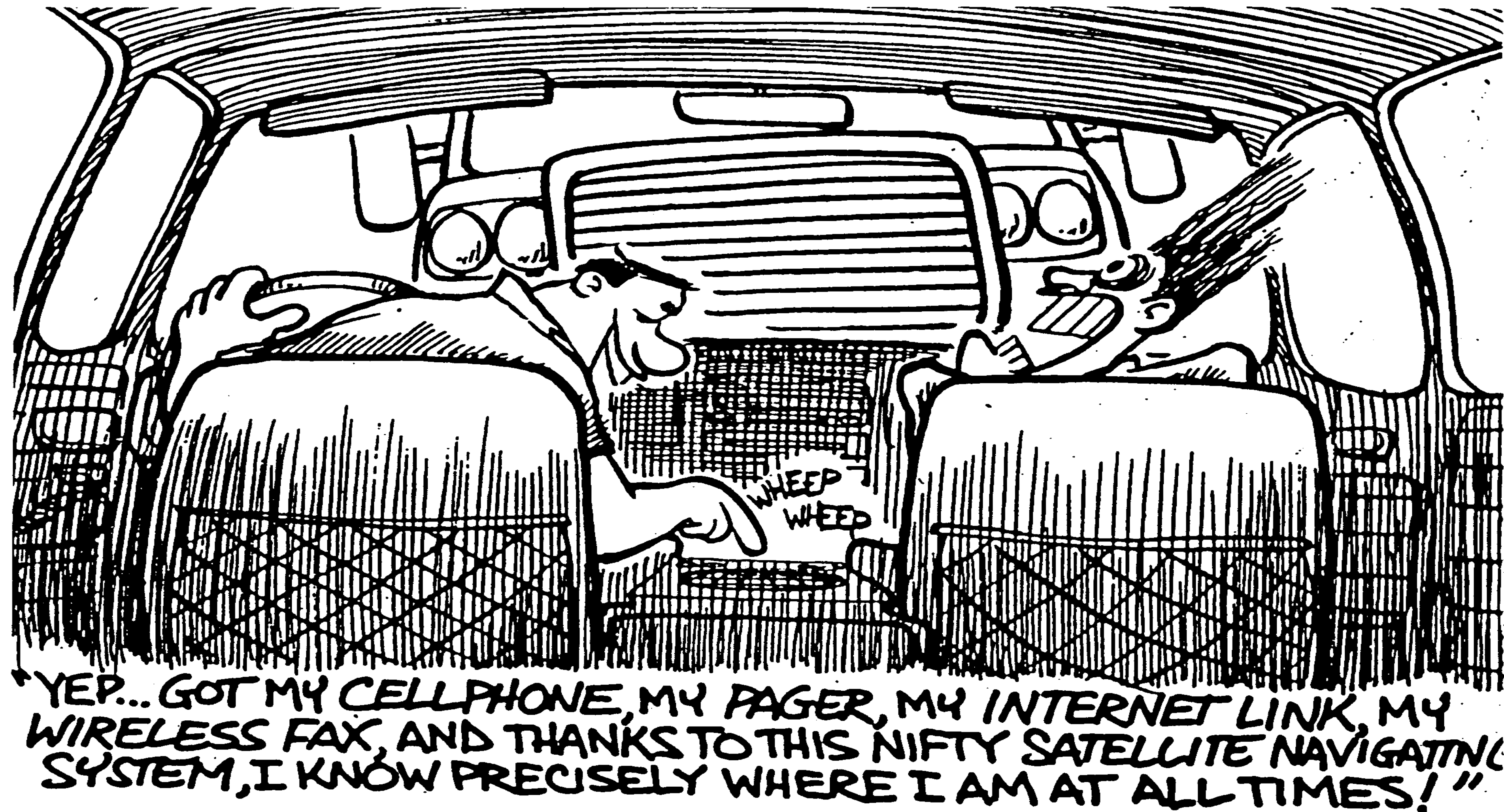


Data is the new oil



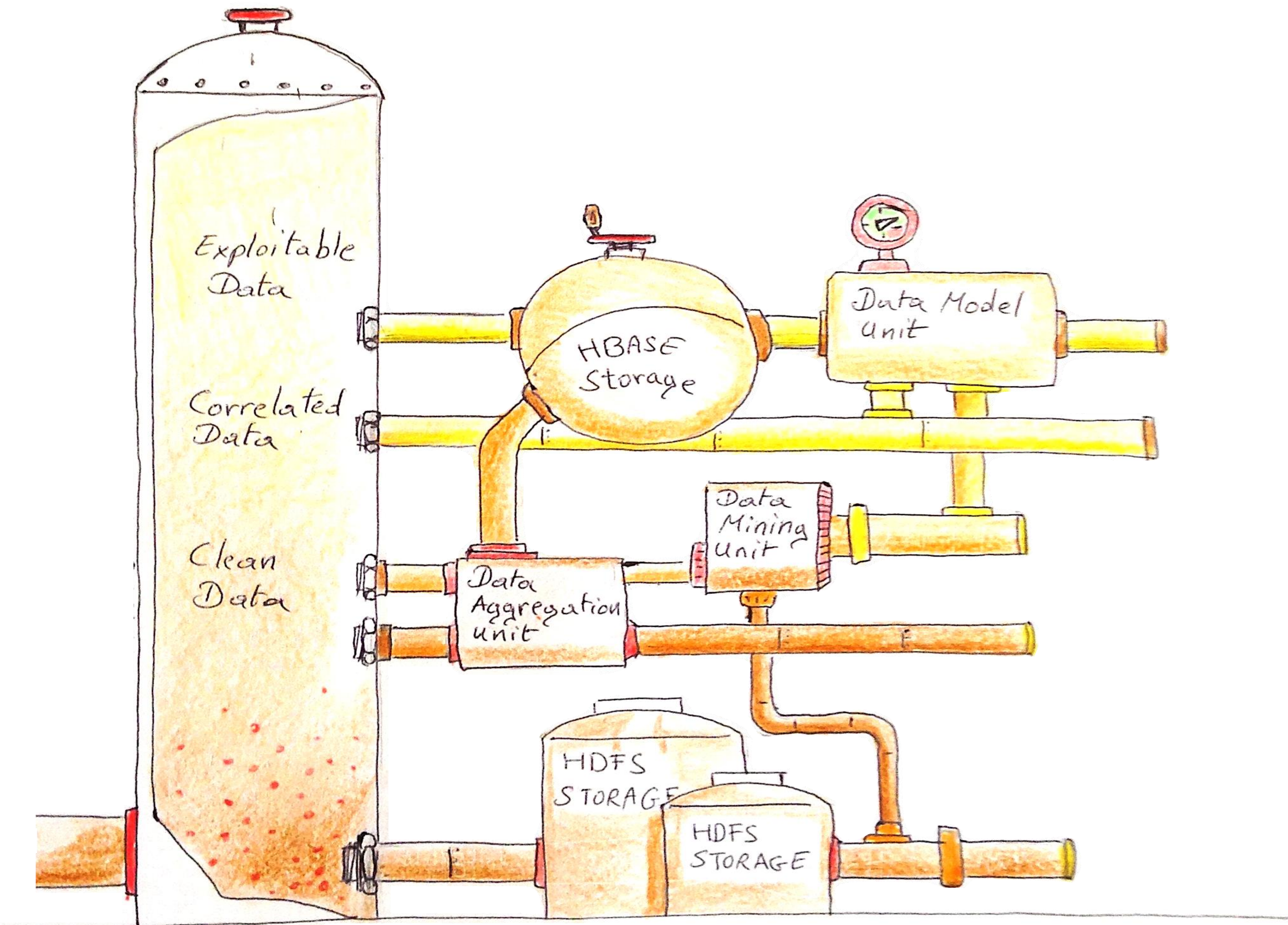
The Economist, May 2017

Big data, Bad data



BY LOWE FOR THE SUN-SENTINEL, FLOR

Like oil, data must be refined

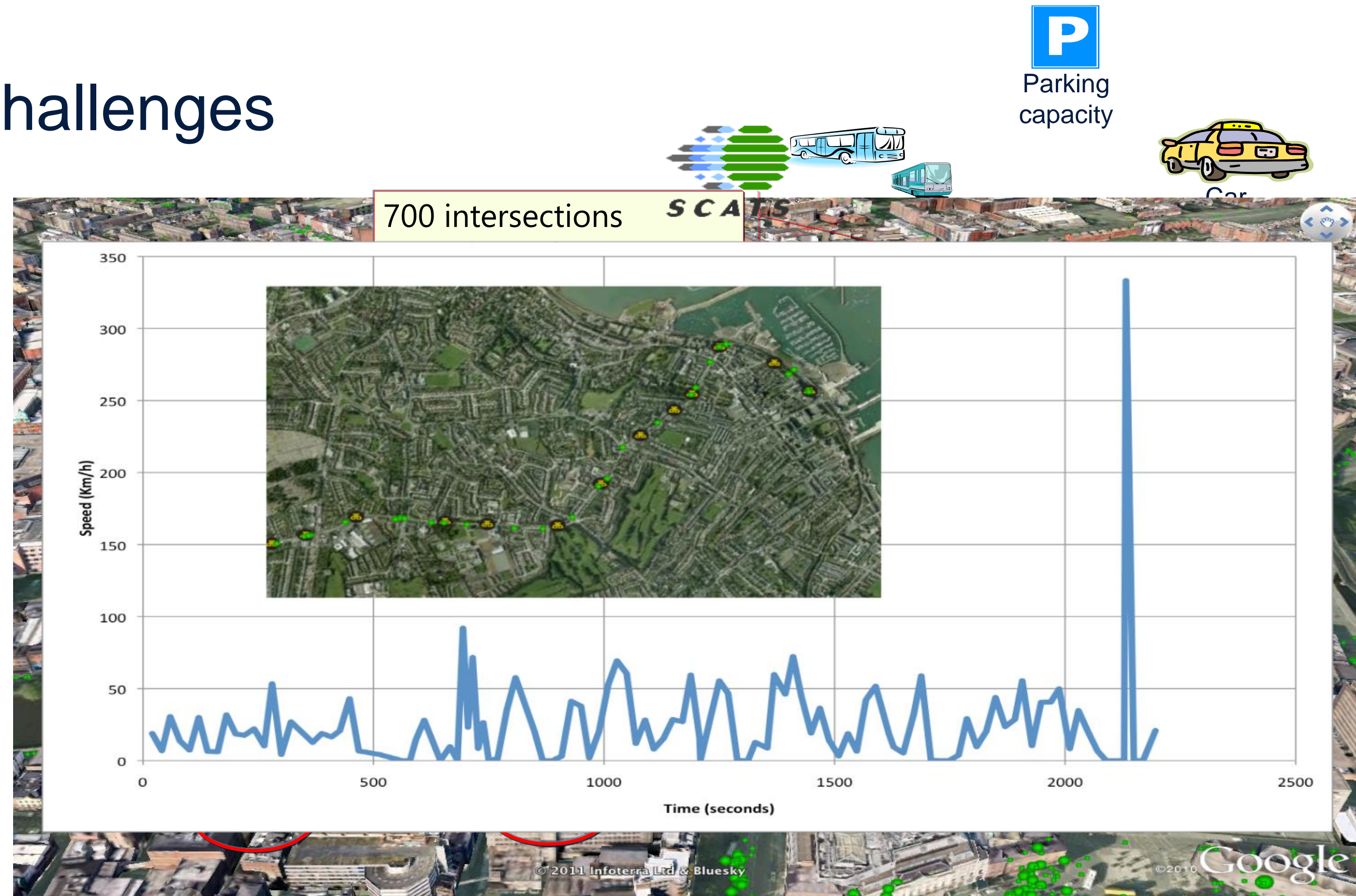


From raw data to information

Real-time tool for situational awareness

- Complex system & analytics challenges

- Data diversity, heterogeneity
- Data accuracy, sparsity
- Data volume



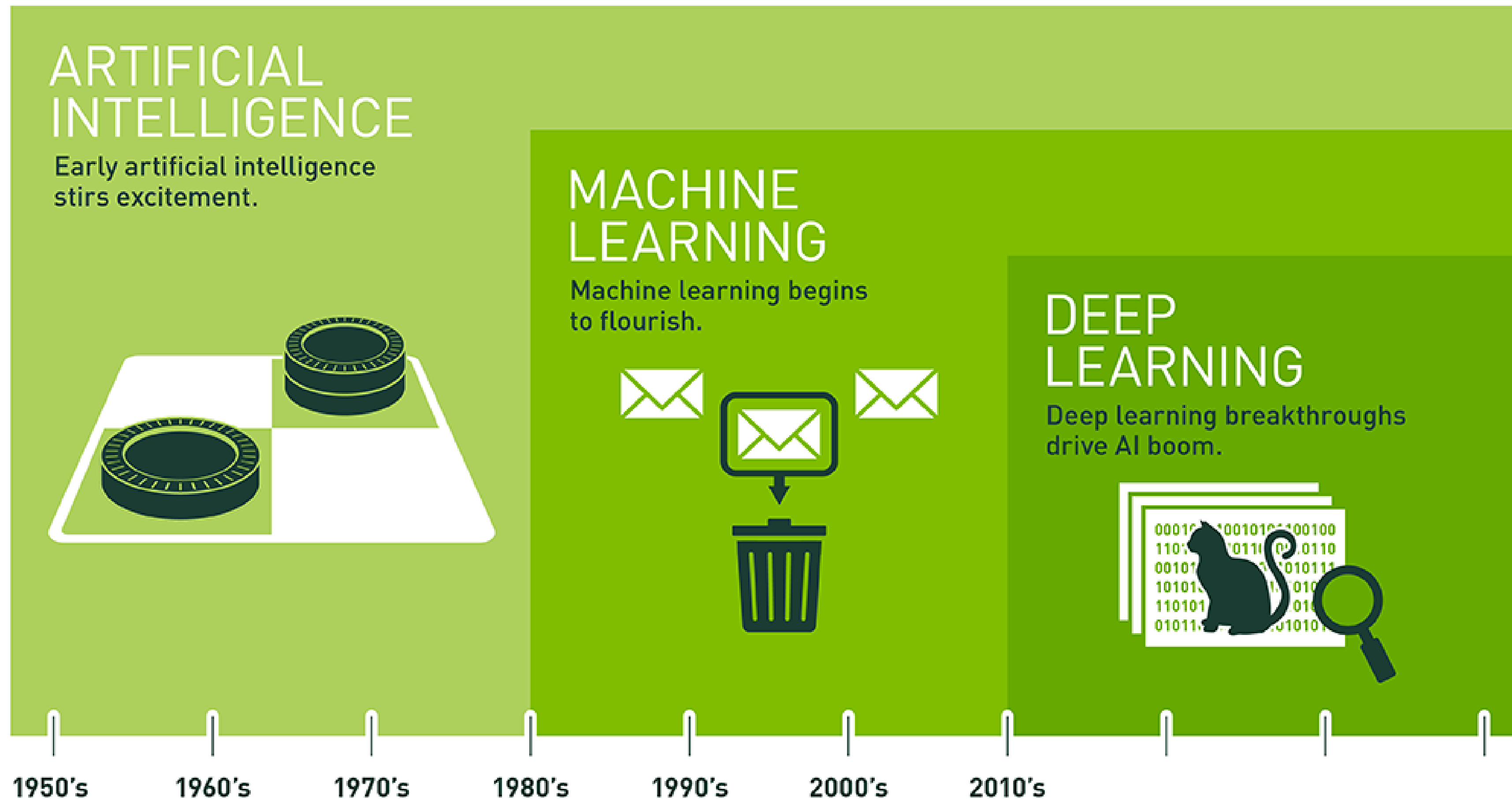
What is Data Science?

Data + Algorithms → Knowledge → Benefits

*Big Data /
Data lake*

*Machine Learning /
"Dumb" A.I.*

ML & AI: A glimpse of history

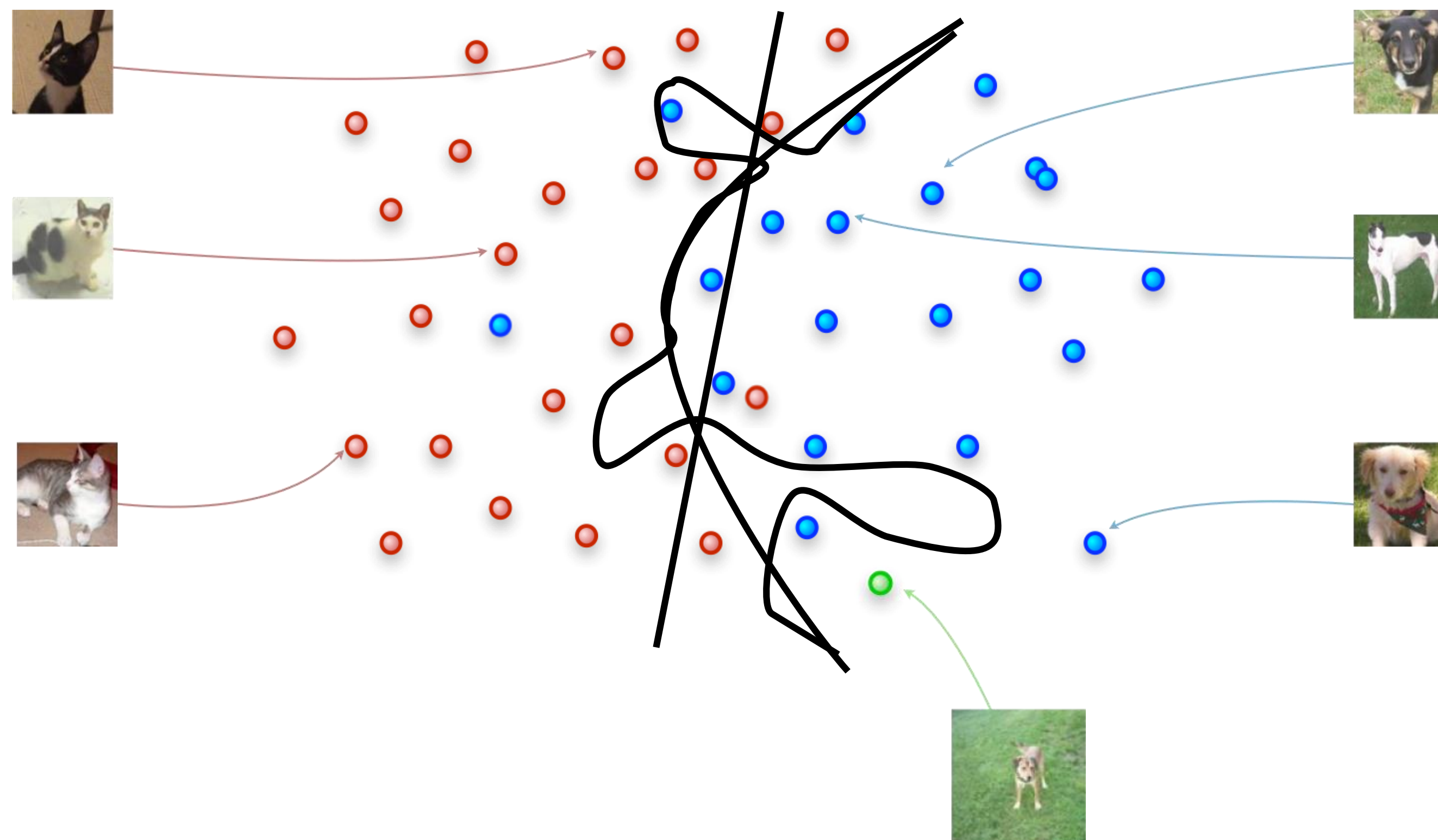


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Machine Learning 101

$$f : \text{cat} \rightarrow \text{Cat}$$

Machine Learning 101



Recent Advances in Machine Learning

Convolutional Neural Networks (CNN)

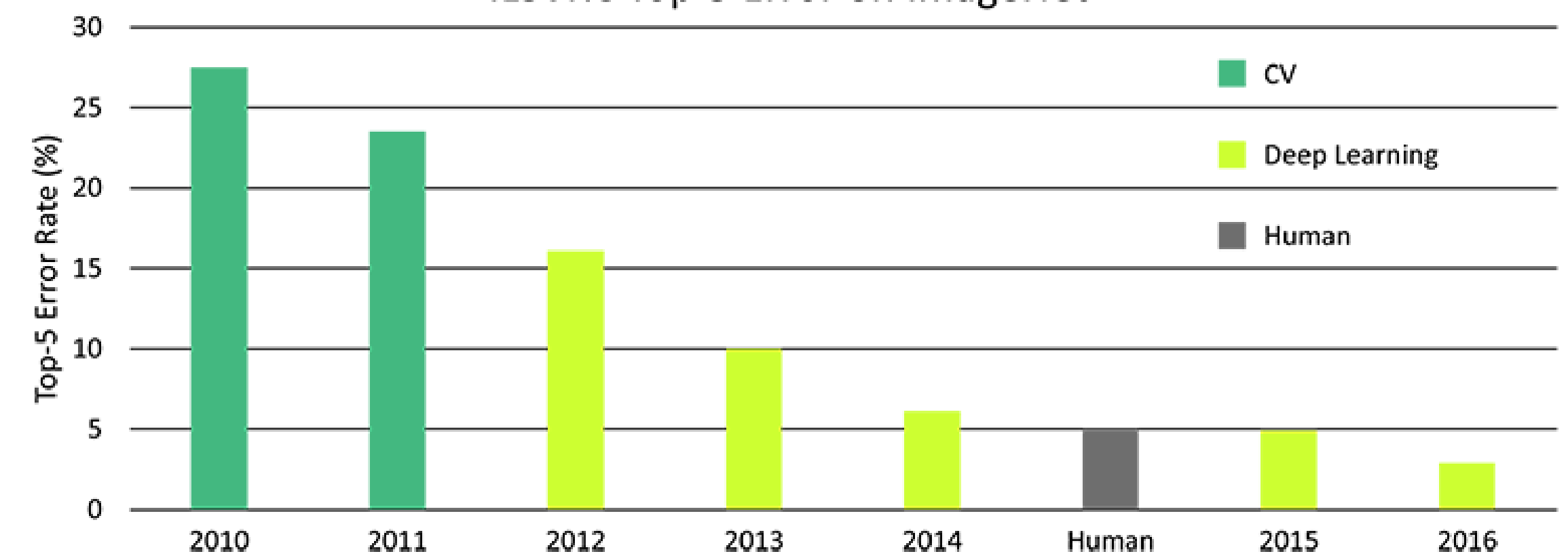
ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.

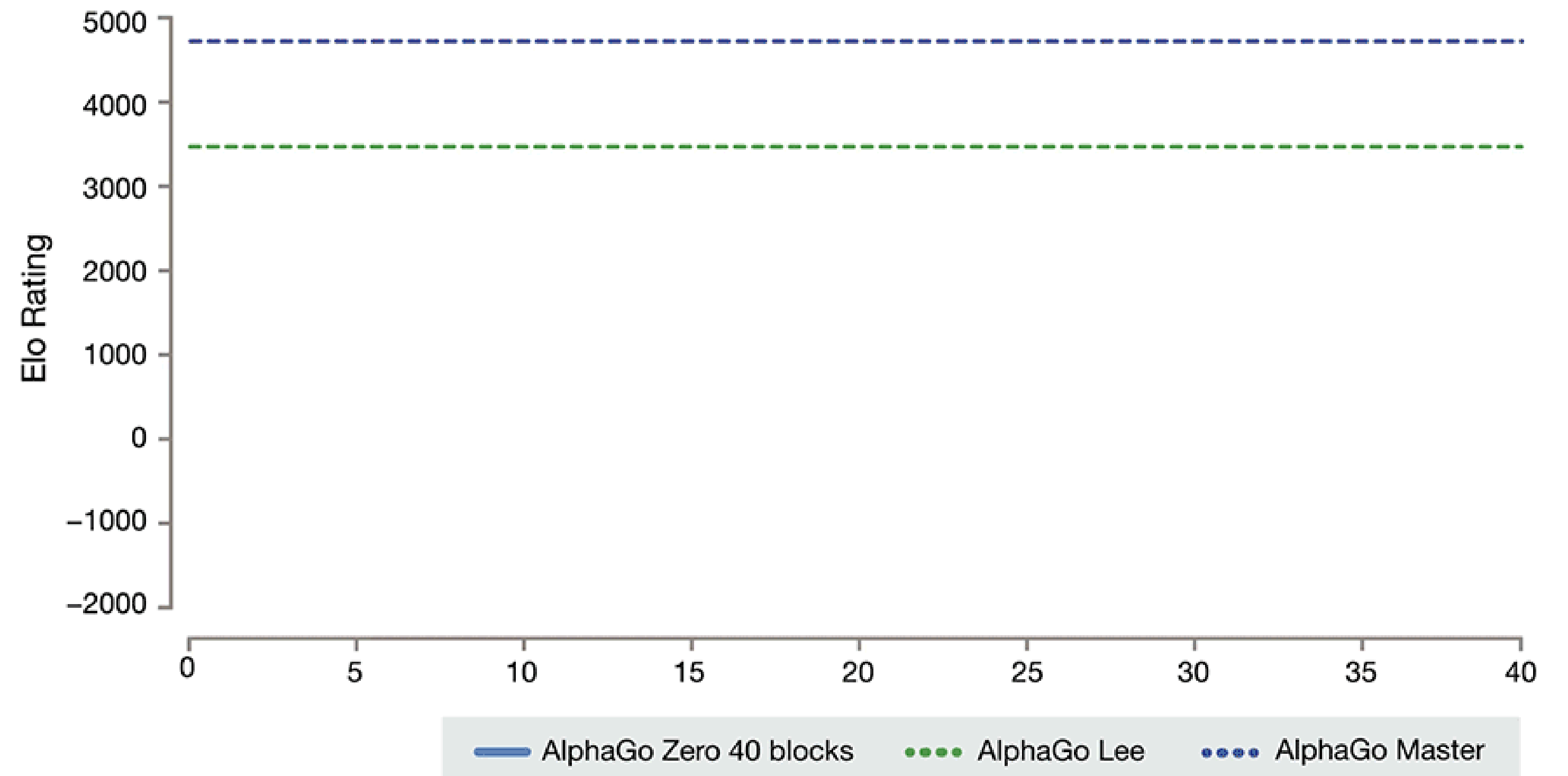


ILSVRC Top 5 Error on ImageNet



Recent Advances in Machine Learning

Deep Reinforcement Learning in AlphaGo Zero



This success relies on...

1. Large dataset of labelled data

2. Good quality data

3. Enough computing power


4. Clear and measurable objectives

An Unexpected Outcome



**It's an Indian
elephant!**

Another Unexpected Outcome



Jonathan Nolis
@skyetetra

Follow

▼

I'm happy to report that hyper-intelligent AI isn't that close: currently Amazon, Microsoft, and Google's cutting-edge APIs all believe "My leg was eaten by a great white shark" is a sentence with a highly positive sentiment.

My leg was eaten by a great white shark

Analyze

67

417

917

Analyzed text	JSON
LANGUAGES:	English (confidence: 100 %)
KEY PHRASES:	great white shark, leg
SENTIMENT:	80 %
LINKED ENTITIES (PREVIEW):	My leg was eaten by a great white shark

A Disturbing Outcome

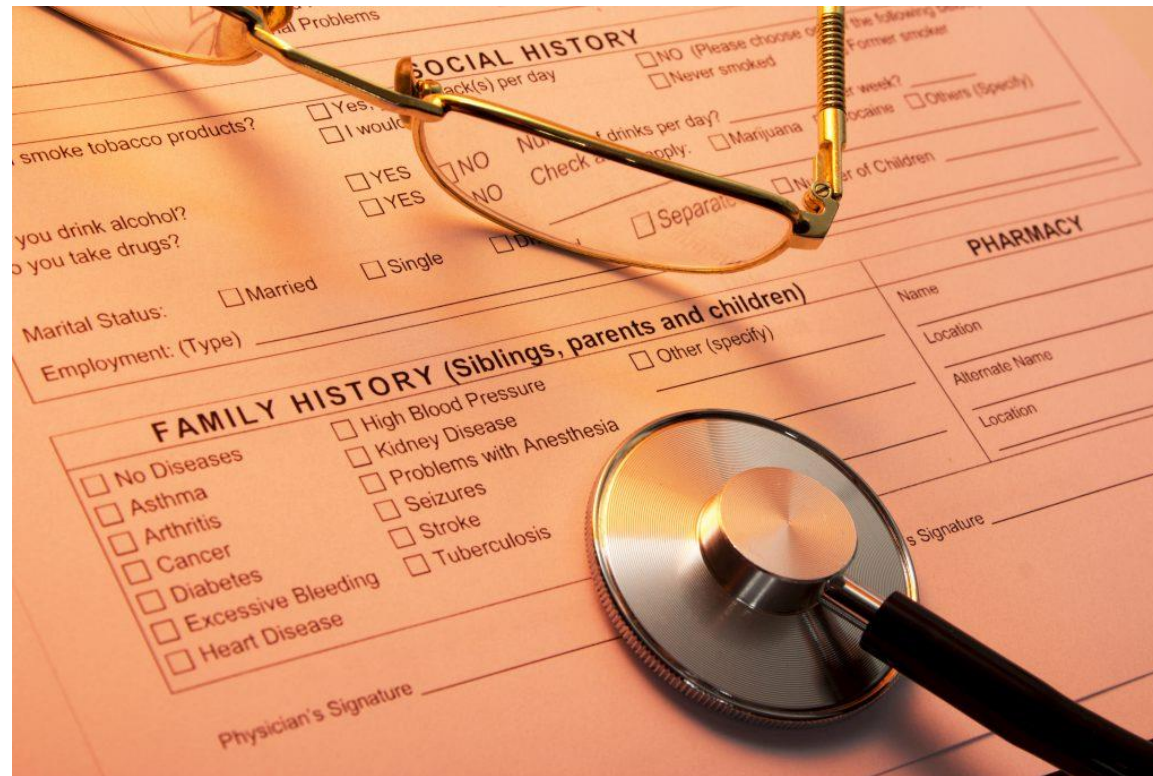


Turning a **STOP** sign into a **60 km/h speed limit**

Unstructured data

$$f : \text{cat} \rightarrow \text{Cat}$$

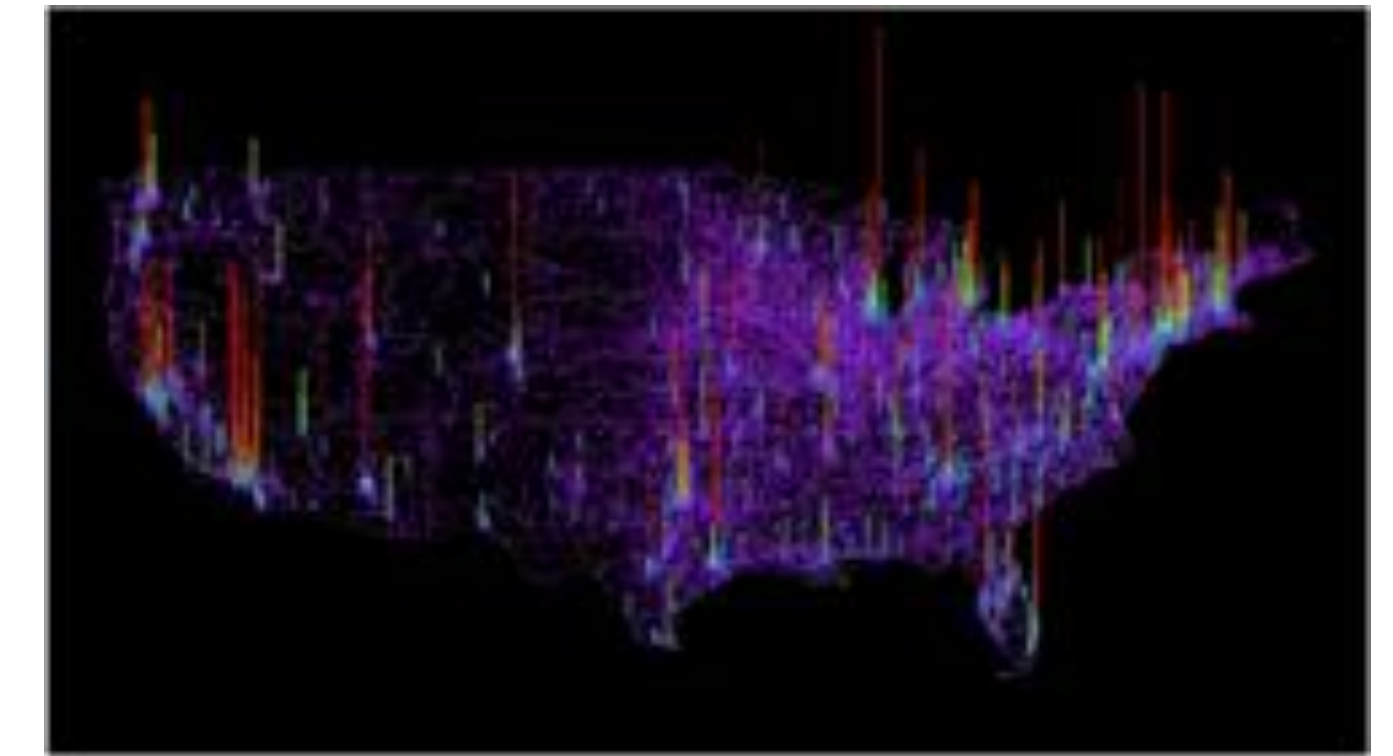
The structured world



Health care



Predictive maintenance



Energy networks



Financial time series

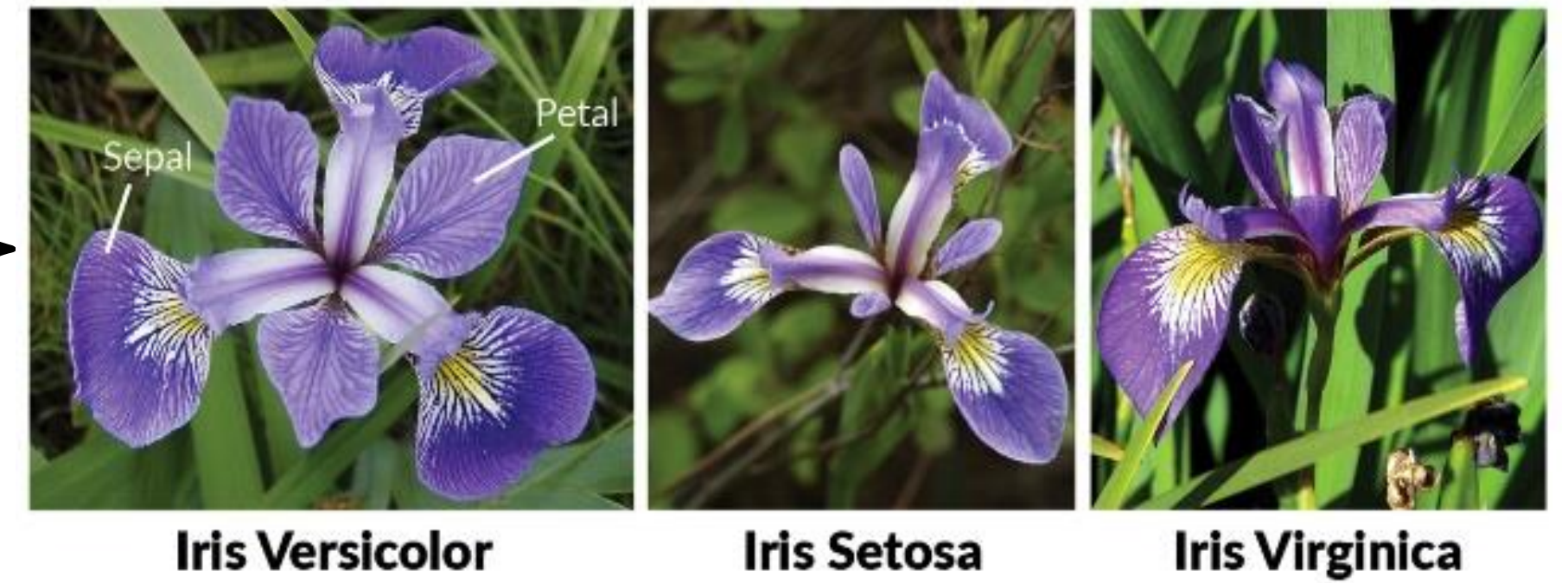
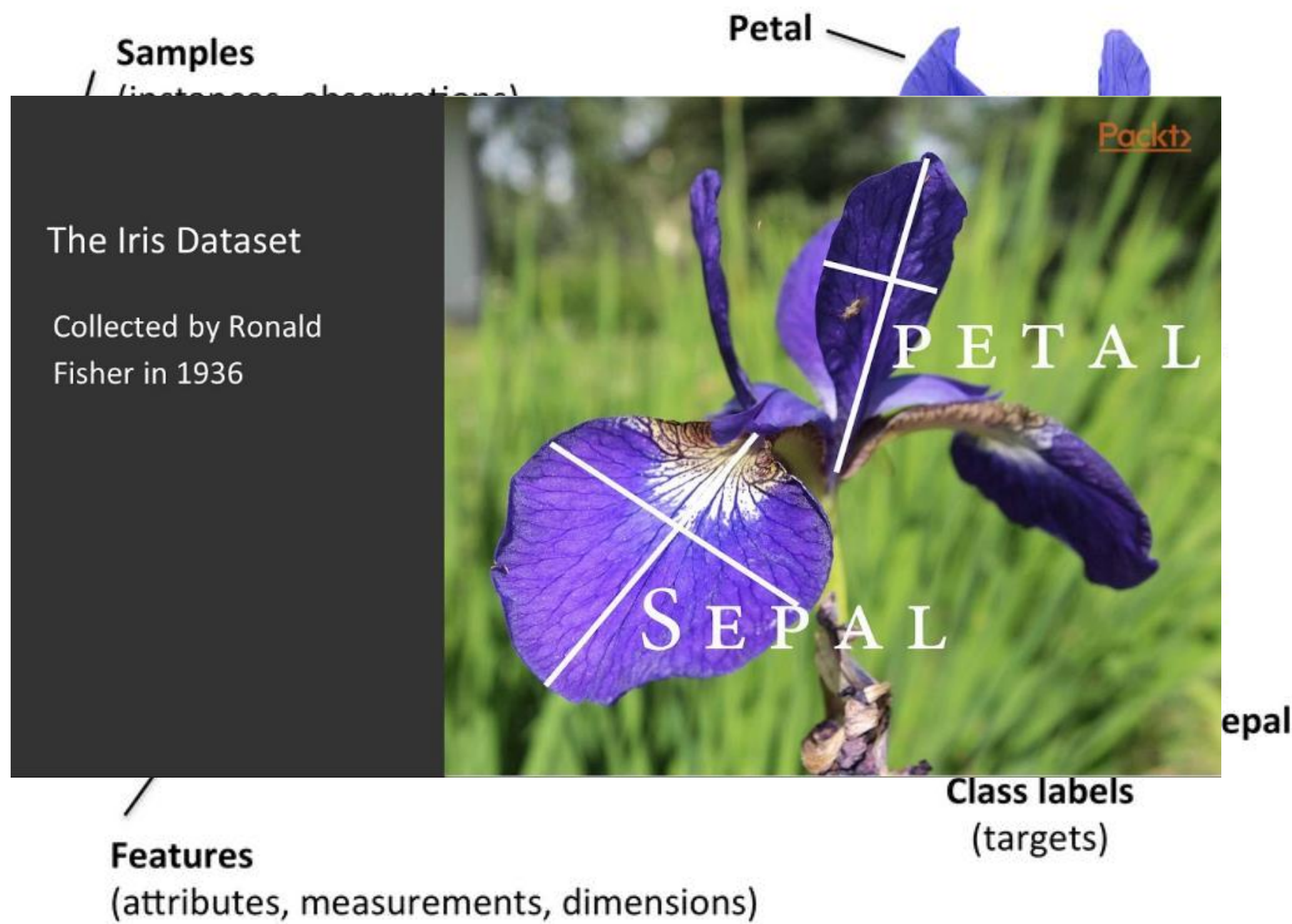
A screenshot of an employee profile page for Nicole Smith. The page includes a profile picture, contact information (email: nicole+icehrm@web-stalk.com), and buttons for 'Edit Info', 'Upload Profile Image', and 'Delete Profile Image'. Below this, there are tabs for 'Basic Information', 'Qualifications', 'Family', and 'Documents'. The 'Basic Information' tab is selected, showing fields for 'Employee Number' (E001), 'NIC', 'Driving License No', 'Other ID', 'Date of Birth' (March 1st, 1984), 'Nationality' (Singaporean), 'Marital Status' (Married), and 'Joined Date' (January 1st, 2014).

Predicting employee attrition

The structured world

f


:



Predicting employee attrition

f

•
•



Nicole Smith
📧 nicole+icehrm@web-stalk.com
[✎ Edit Info](#) [📷 Upload Profile Image](#) [✕ Delete Profile Image](#)

Employee Number

NIC

E001

Basic Information

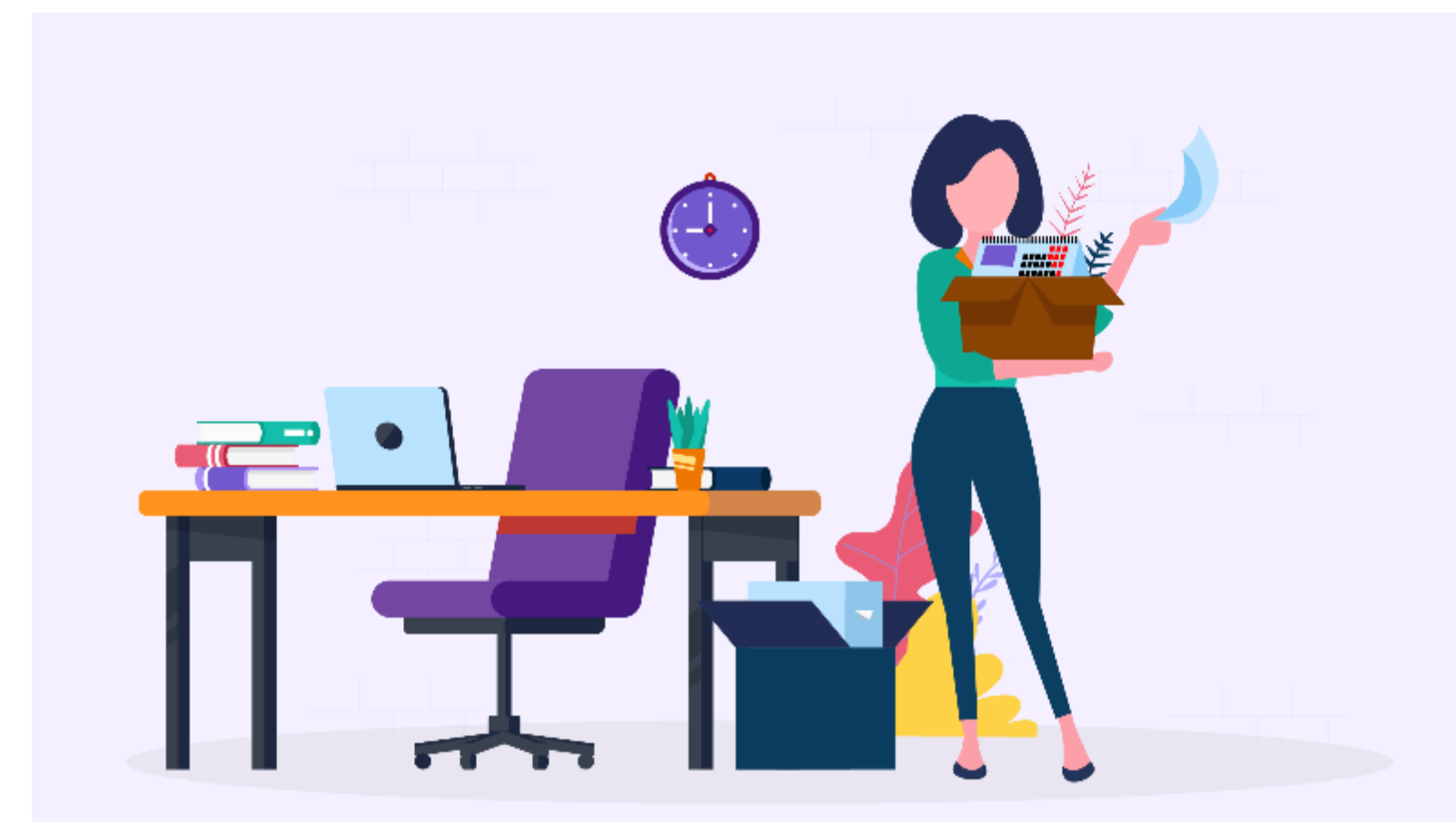
Qualifications

Family

Documents

Personal Information

Driving License No	Other ID	Date of Birth
Nationality	Marital Status	Joined Date
Singaporean	Married	March 1st, 1984
		January 1st, 2014



Obstacles to a wider adoption

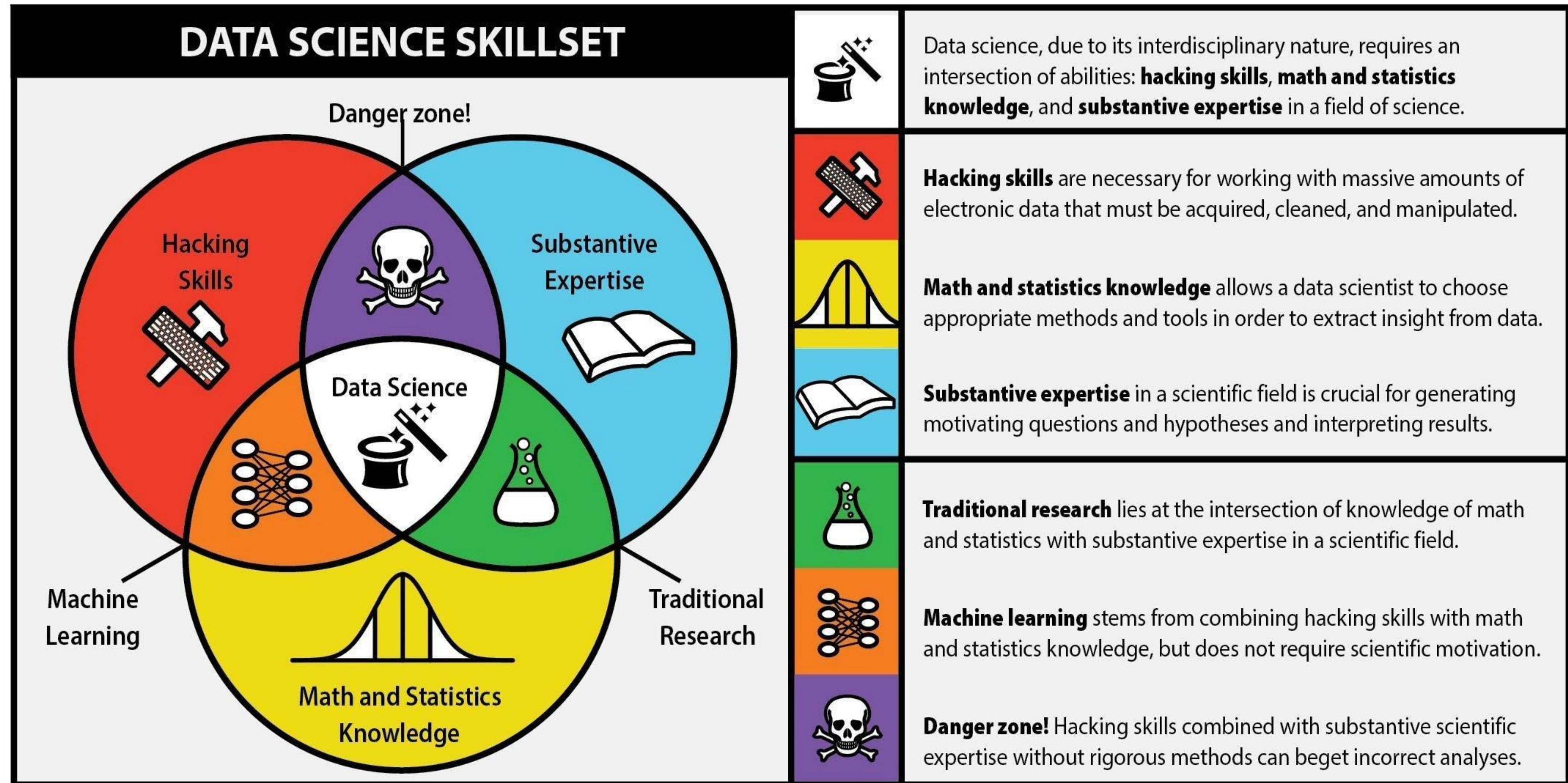
1. ~~Large dataset of labelled data~~ -> **Labelling is expensive**

2. ~~Good quality data~~ -> **Data is usually missing/Increased uncertainty**

3. ~~Clear and measurable objectives~~ -> **Knowledge discovery/causality**

4. **Lack of interpretability/lack of trust**

From Machine Learning to Data Science



What is Data Science?

Domain expertise

Data + Algorithms → Knowledge → Benefits

*Big Data /
Data lake*

*Machine Learning /
"Dumb" A.I.*

Data Scientist skills

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



IN GOD WE TRUST.



ALL OTHERS MUST BRING DATA.

– W. EDWARDS DEMING, STATISTICIAN, PROFESSOR, AUTHOR