Ollscoil na Gaillimhe
University of Galway

CT4101
Machine Learning

Dr. Frank Glavin
Room 404, Computer Science Building
Frank.Glavin@University*of*Galway.ie

School of Computer Science

University *of* Galway.ie

Overview

1. Learning objectives and overview
2. What is Machine Learning?
3. Types of Machine Learning task
4. Overview of Data Mining
5. Applications of Machine Learning

Learning objectives for this section

Having completed this section successfully, you will be able to …

1. Discuss definitions of Machine Learning
2. Describe what major categories of ML task entail: classification, regression, clustering, relationship discovery and reinforcement learning
3. Discuss the relationship with Data Mining
4. Explain the Data Mining process
5. Consider current and future applications of Machine Learning and Data Mining

Resources

Course slides:

- Necessary but insufficient!

Recommended books:

- List available on Blackboard

Others:

- Andrew Ng's Coursera Machine Learning Course
- Sebastian Thrun's Udacity AI Course
- Contributions welcome!
- If you find useful links, email them to me at frank.glavin@universityofgalway.ie

What is Machine Learning? [1]

Samuel, 1959:

"*Field of study that gives computers the ability to learn without being explicitly programmed*"

Witten & Frank, 1999:

Learning is changing behaviour in a way that makes performance better in the future

Arthur Samuel, 1901-1990

Image source: http://www.computer.org/portal/web/awards/cp-samuel

What is Machine Learning? [2]

Mitchell, 1997:

- Improvement with experience at some task
- A well-defined ML problem:
  - Improve over task *T*
  - wrt performance measure *P*
  - based on experience *E*
- For draughts/checkers example, what are T, P, E?

Other possible definitions

- Philosophical and psychological considerations …
- Relationship to Artificial Intelligence generally …
- Artificial Intelligence ≠ Machine Learning ≠ Deep Learning
- **Artificial Intelligence ⊋ Machine Learning ⊋ Deep Learning**

http://www.cs.cmu.edu/~tom/

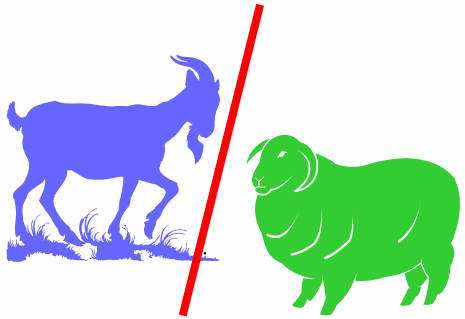
What is Machine Learning? [3]

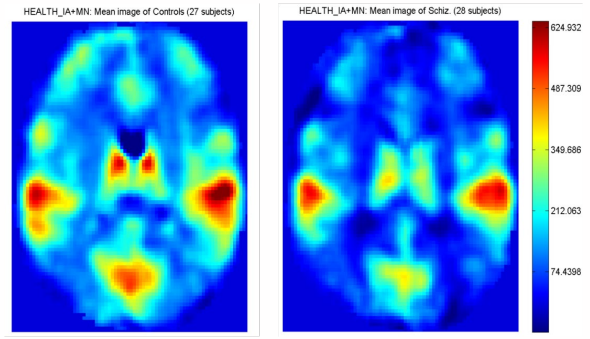
Machine Learning Techniques

Supervised learning
Unsupervised learning
Semi-supervised learning
Reinforcement learning

Major Types of Task [1]

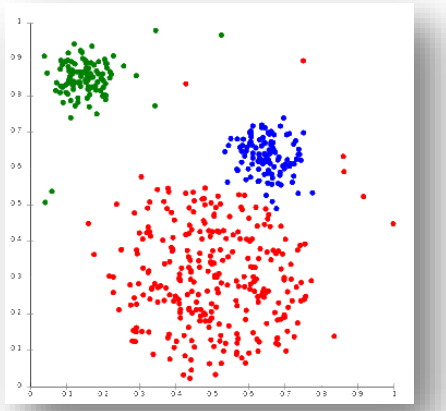
1. **Classification**





2. **Regression**

3. **Clustering**

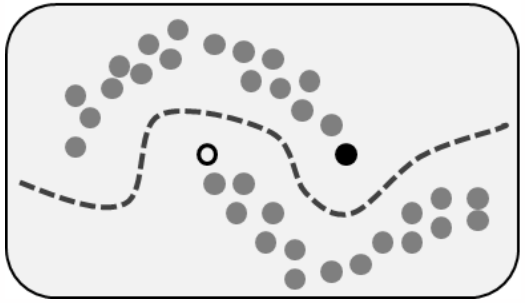


Major Types of Task [2]

4. **Co-Training**

5. **Relationship Discovery**

6. **Reinforcement Learning**



beer <=> diapers

Techniques for these Tasks

Classification
Decision trees, SVMs

Regression
Linear Regression, Neural nets; k-NN
(good for Classification too)

Supervised

Clustering
k-Means, EM-clustering

Relationship Discovery
Association Rules; Bayesian nets

Unsupervised

Learning From Part-Labelled Data
Co-Training; Transductive Learning
[Combines ideas from clustering & classification]

Semi-supervised

Reinforcement Learning
Q-Learning, SARSA

Reward-based

What do these have in common?

In all cases, machine searches for a hypothesis that best describes the data presented to it

Choices to be made:

- How is hypothesis expressed?
  mathematical equation, logic rules, diagrammatic form, table, parameters of a model (e.g. weights of an ANN), …
- How is search carried out?
  systematic (breadth-first or depth-first),
  heuristic (most promising first), …
- How do we measure quality of hypothesis?
- What is appropriate format for data?
- How much data is required?

What else to we need to know about?

To apply ML:

- How to formulate a problem
- How to prepare the data
- How to select an appropriate algorithm
- How to interpret the results

To evaluate results and compare methods:

- Separation between training, testing & validation
- Performance measures:
  simple metrics, statistical tests, graphical methods
- To improve performance
  Ensemble methods
- Theoretical bounds on performance

Data Mining: What's the Link?

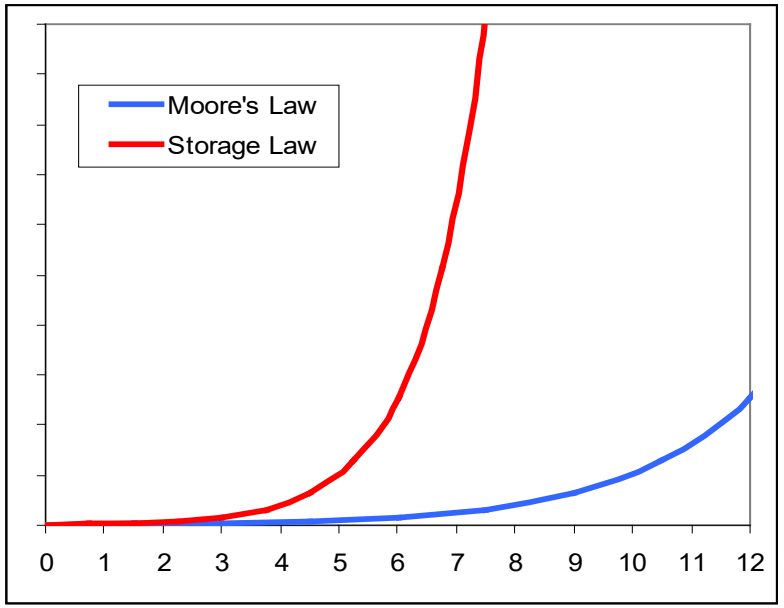
Data Mining:

Extract interesting knowledge from large unstructured datasets
non-obvious / comprehensible / meaningful / useful

Storage Law (Fayyad & Uthurusamy, Comms.ACM 2002)

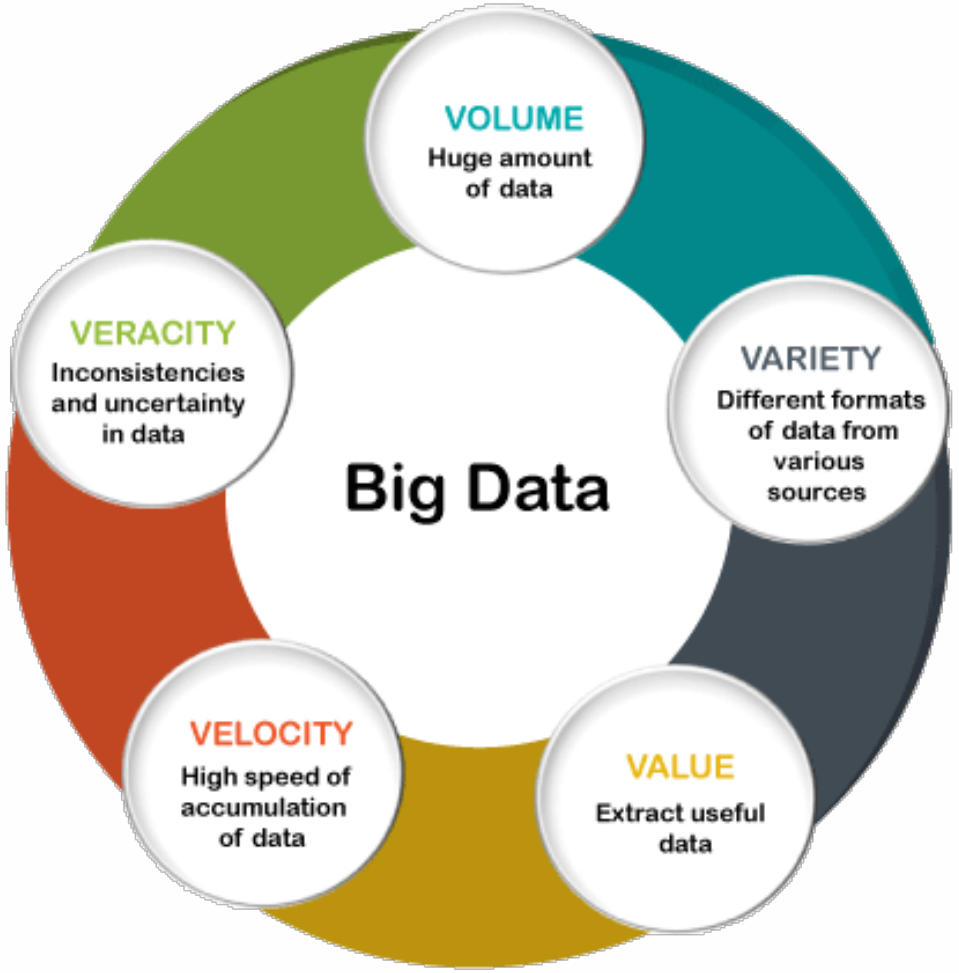
Storage capacity doubling every year
Faster than Moore's law
Result: write-only "data tombs"

Developments in ML essential
to be able to process
and exploit this lost data



Big Data

- Data sets of scale and complexity such that they can be difficult to process using current standard methods
  - Standard DB tools & data management apps
  - Moving target



Source

Big Data

Data scale dimensions (One or more of “3 Vs”):

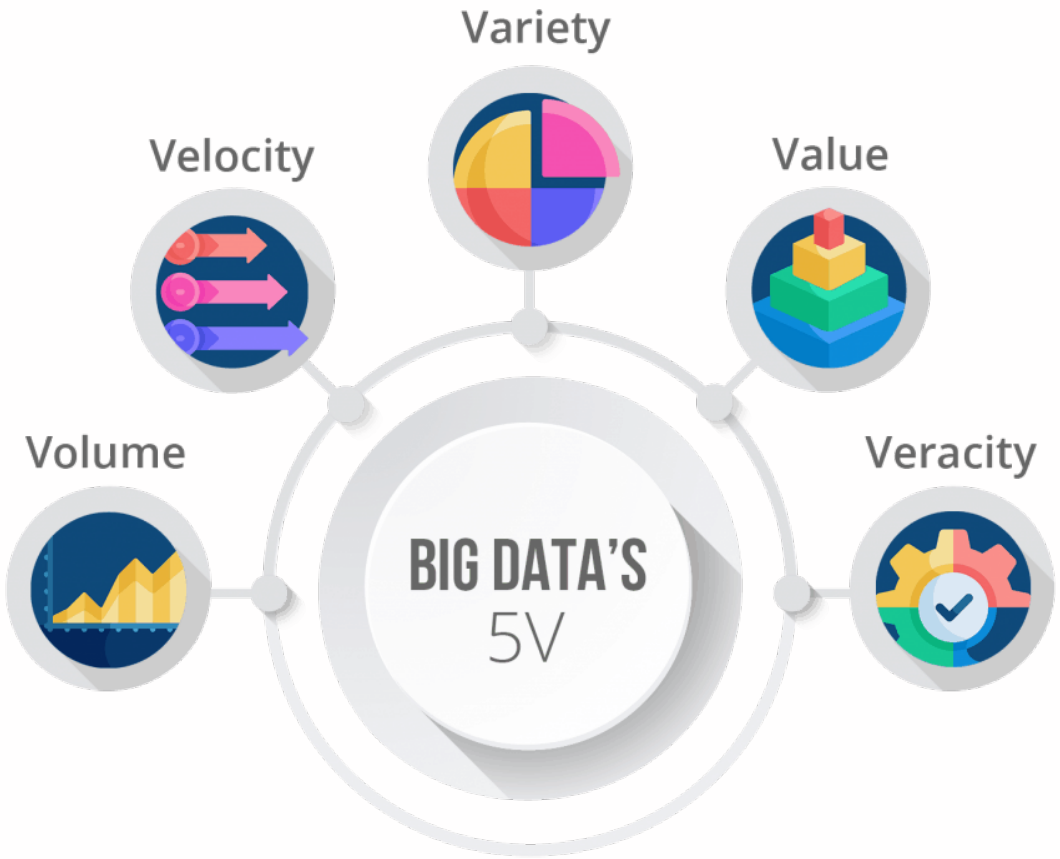
- **Volume**: terabytes and up
- **Velocity**: from batch to streaming data
- **Variety**: numeric, video, sensor, unstructured text …

Fashionable to add others that are not key …

- Veracity: quality & uncertainty associated with items
- Variability: change / inconsistency over time
- Value: for the organization

Key techniques:

- Sampling; inductive learning; clustering; associations
- Distributed programming methods

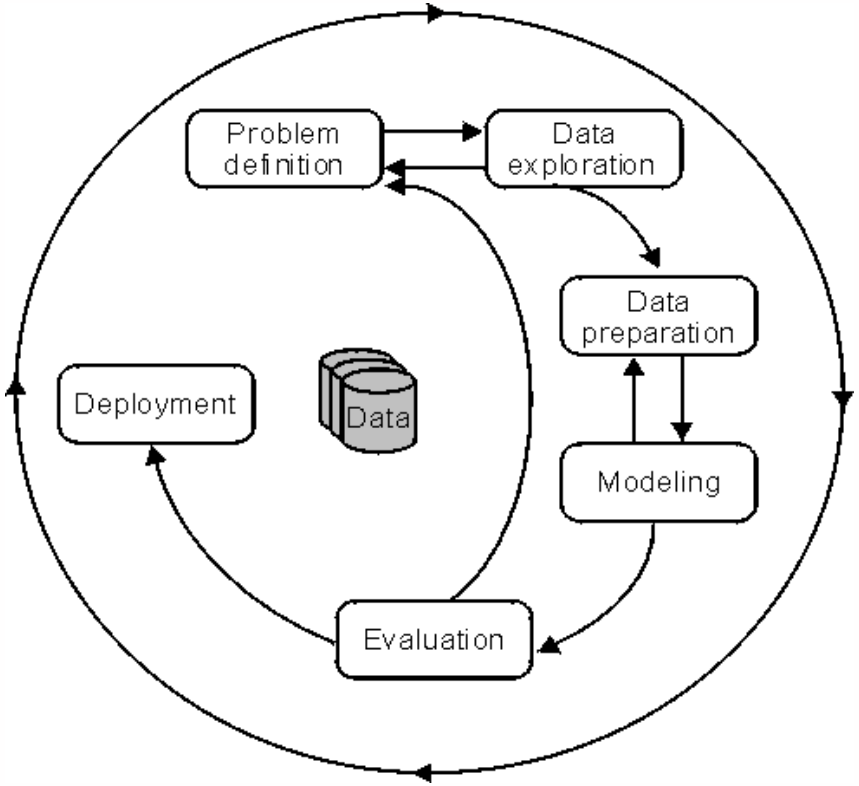


CRISP-DM Data Mining Process

Problem Definition
Data Exploration
Data Preparation
Modelling
Evaluation
Deployment

Cross Industry Standard Process for
Data Mining (CRISP-DM) process model

This link gives a summary of the main steps in CRISP-DM
https://www.sv-europe.com/crisp-dm-methodology/



Current & Emerging Applications

Any ideas?

What companies use Machine Learning?

Ollscoil na Gaillimhe
University of Galway

School of Computer Science

Users of ML & DM

amazon

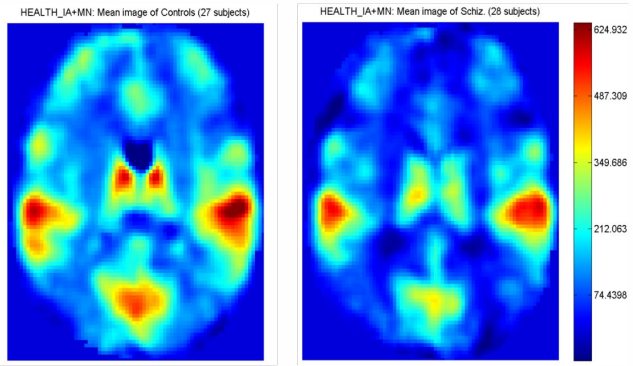
NETFLIX

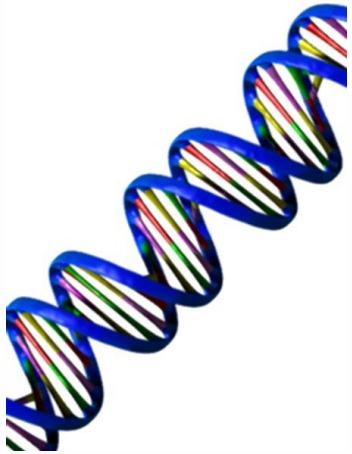
Microsoft

YAHOO!

facebook

Google





High-Profile Examples …

Forbes · New Posts +4 posts this hour · Most Popular Most Disliked Athletes

Kashmir Hill, Forbes Staff

Welcome to The Not-So Private Parts where technology & privacy collide

+ Follow (1,178)

TECH | 2/16/2012 @ 11:02AM | 1,930,513 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Forbes, 16 Feb 2012

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

House of Cards (BBC, 1990)

⇔ ★ ★ ★ ★ ★

⇔ Kevin Spacey (Actor)

⇔ David Fincher (Dir.)

How Netflix is turning viewers into puppets

"House of Cards" gives viewers exactly what Big Data says we want. This won't end well

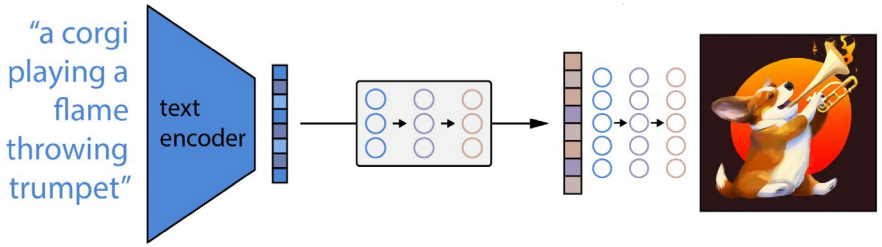
BY ANDREW LEONARD

Salon, 1 Feb 2013
https://www.salon.com/2013/02/01/how_netflix_is_turning_viewers_into_puppets/

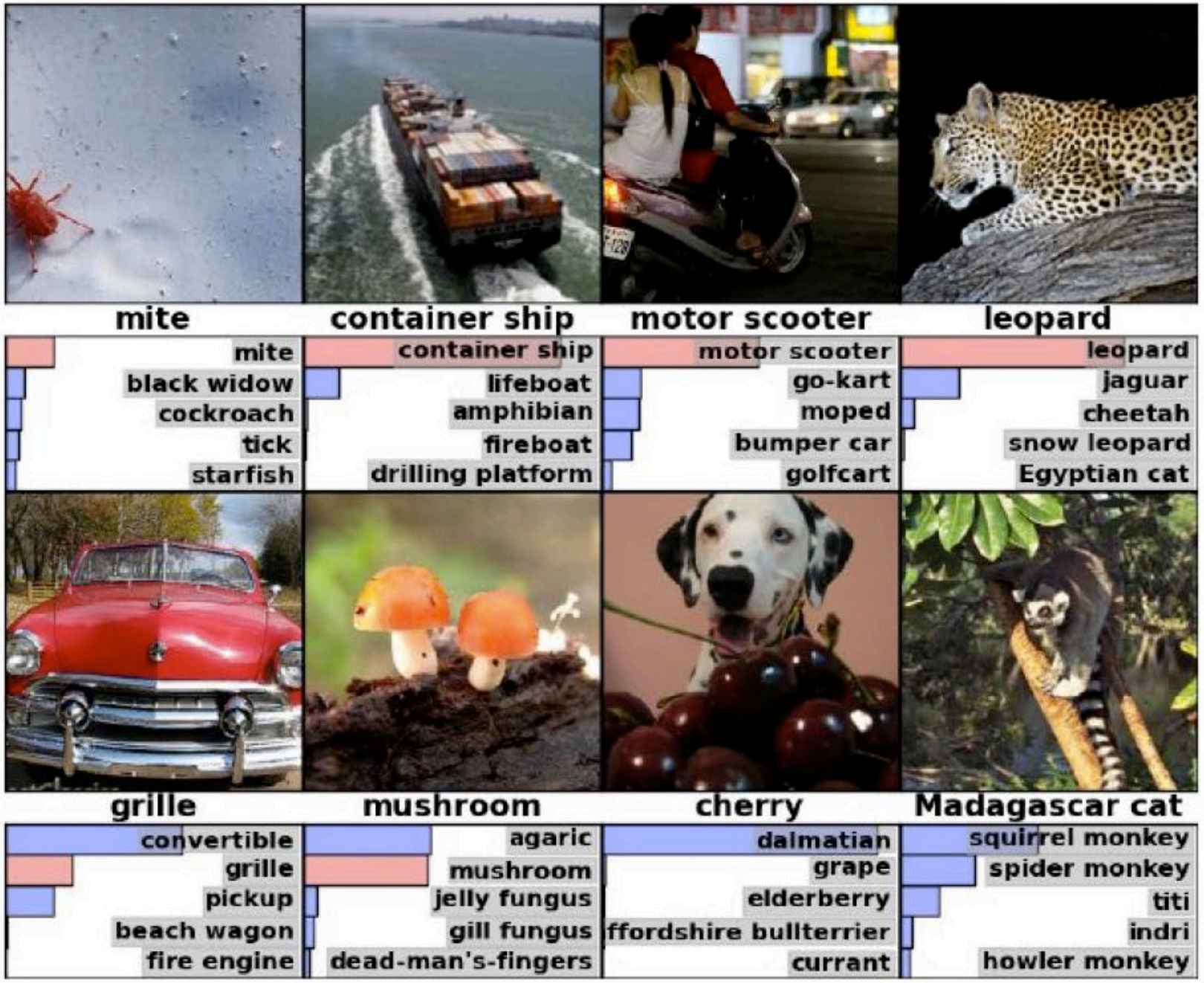
DALL-E 2

https://openai.com/dall-e-2/

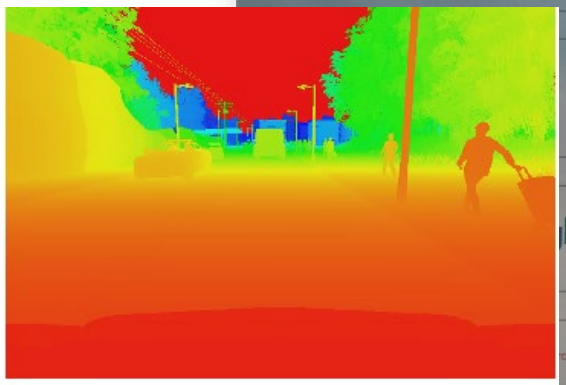
https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

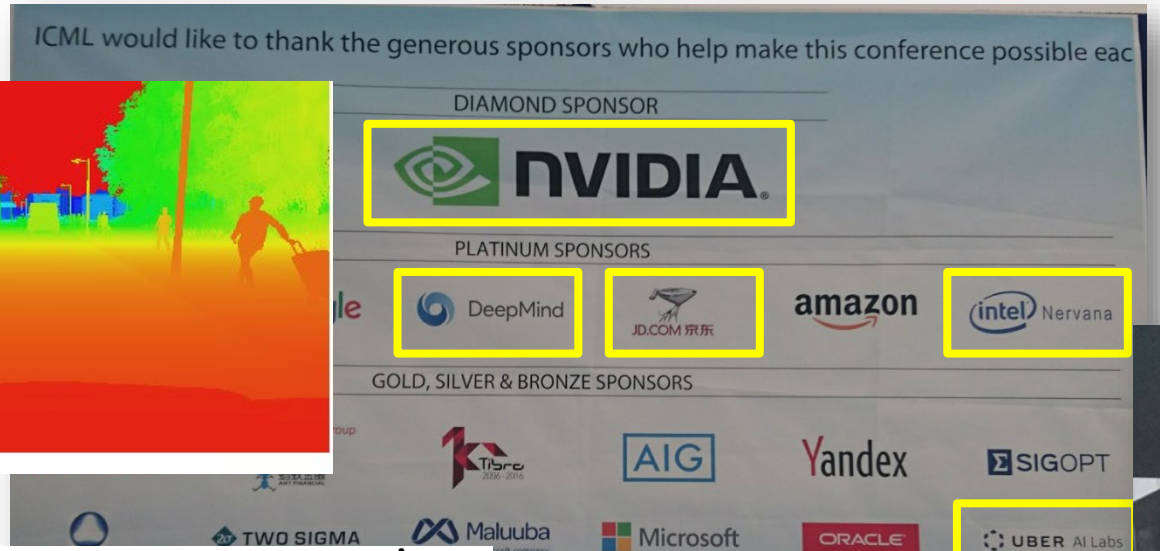


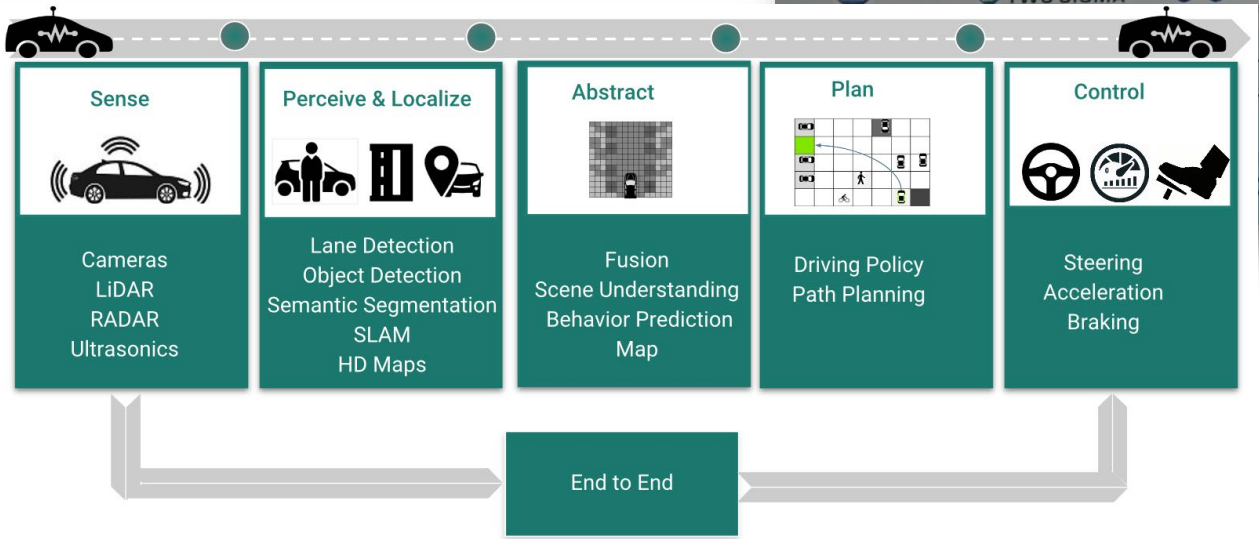
Deep Learning for Object Recognition: Hinton & colleagues, NIPS 2012



AI/ML for Autonomous Vehicles





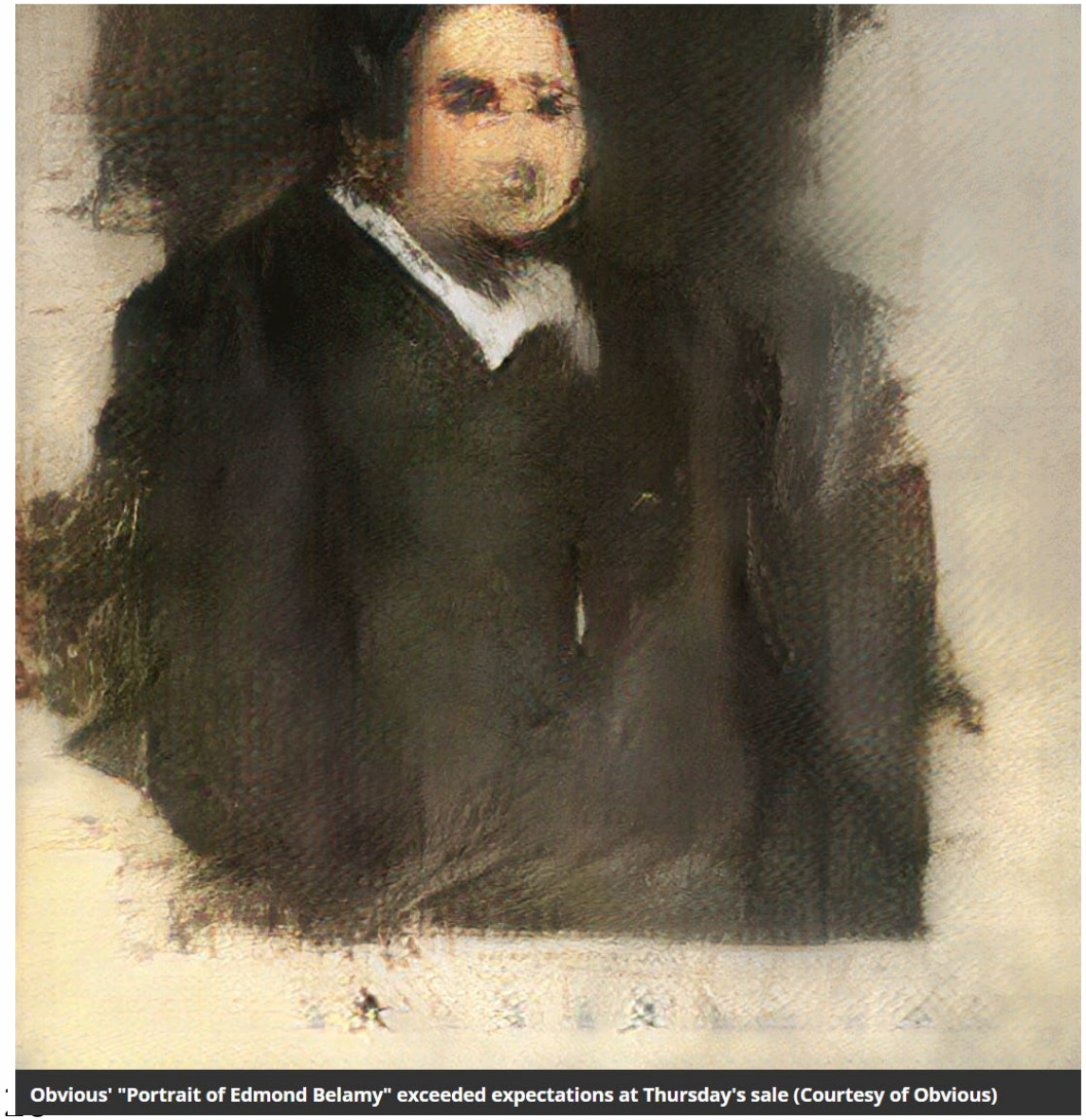


Science

Learning from Experience

elf

Generative Adversarial Networks



Obvious' "Portrait of Edmond Belamy" exceeded expectations at Thursday's sale (Courtesy of Obvious)

SMARTNEWS *Keeping you current*

Christie's Is First to Sell Art Made by Artificial Intelligence, But What Does That Mean?

Paris-based art collective Obvious' 'Portrait of Edmond Belamy' sold for $432,500, nearly 45 times its initial estimate

October 2018

https://www.smithsonianmag.com/smart-news/christies-first-sell-art-made-artificial-intelligence-what-does-mean-180970642/

https://medium.com/syncedreview/nvidia-open-sources-hyper-realistic-face-generator-stylegan-f346e1a73826
https://www.theverge.com/2019/3/19/18272602/ai-art-generation-gan-nvidia-doodle-landscapes

Generative Adversarial Networks

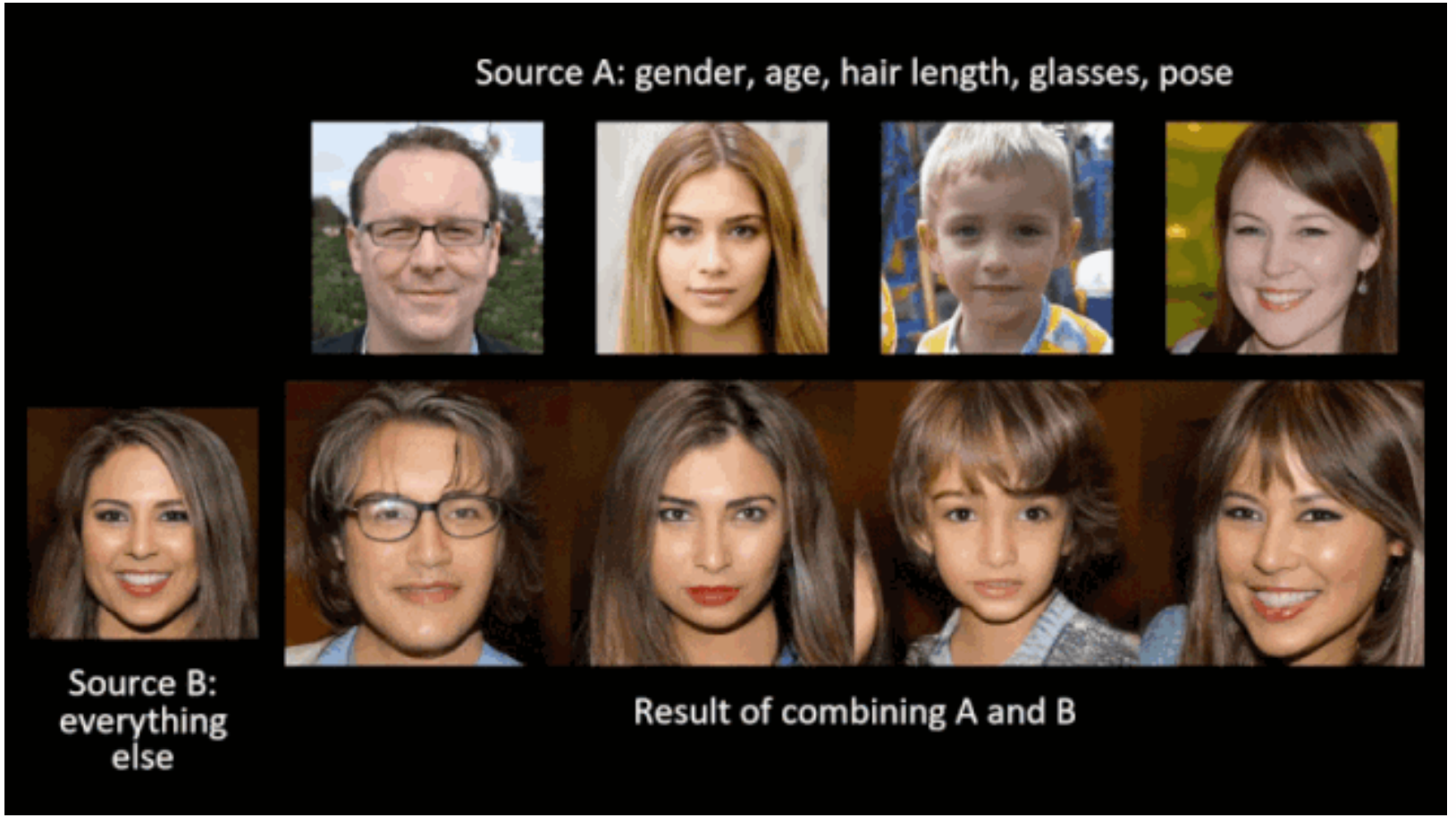
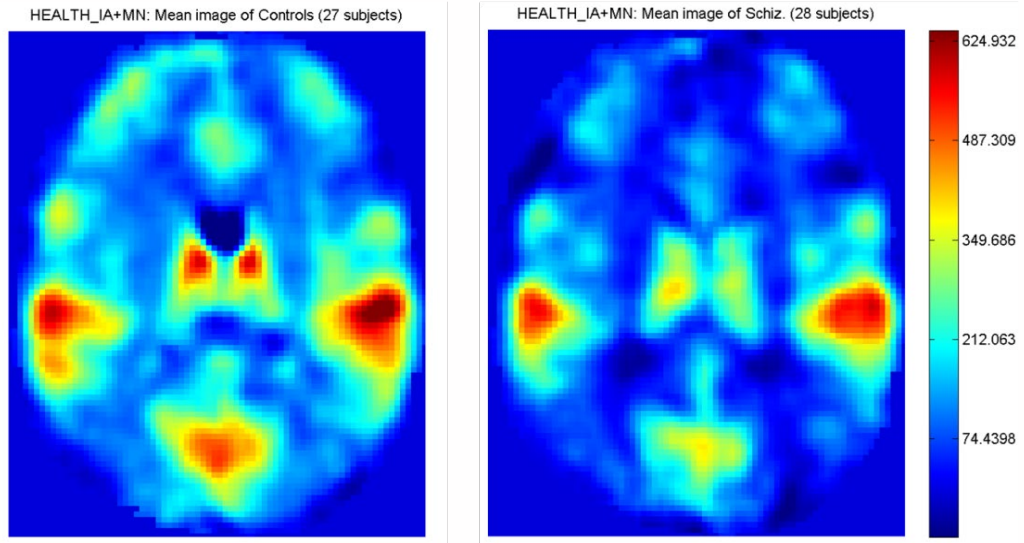
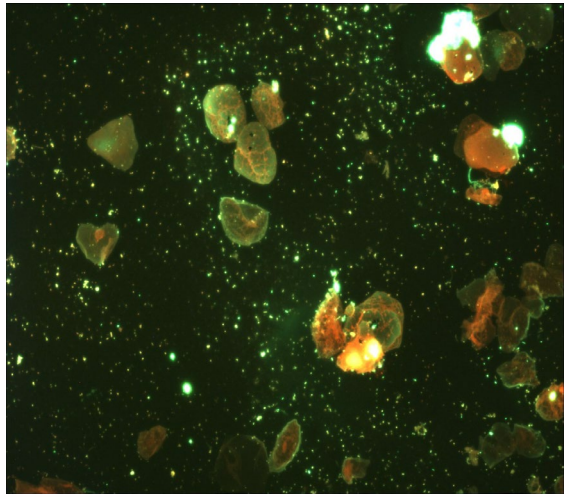


Image & Sensor Data Mining

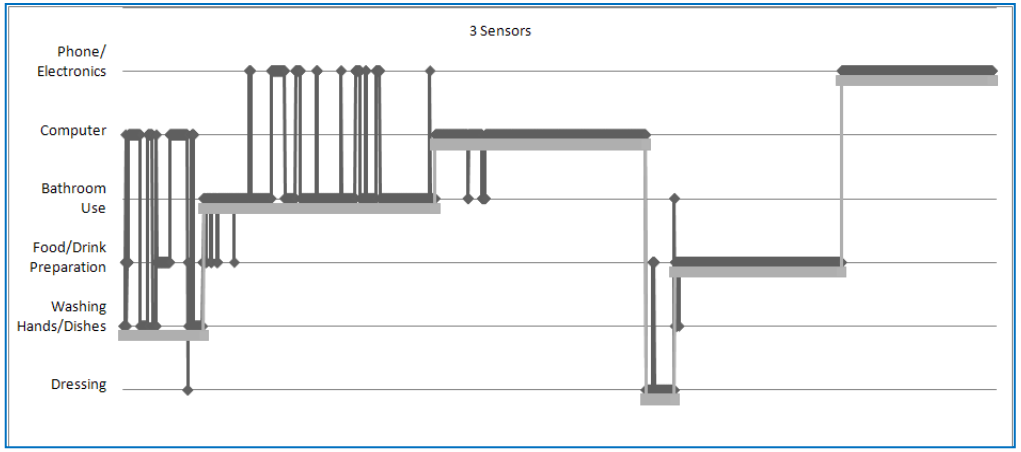
UC Irvine / NIH BIRN collaboration:
Using fMRI to distinguish subjects with Schizophrenia from controls





Analysing microscope images of sputum to screen for TB:
Image processing, ML,
Sequential statistics

Identify Activities of Daily Living from sensors:
Ensemble DTW Classifier

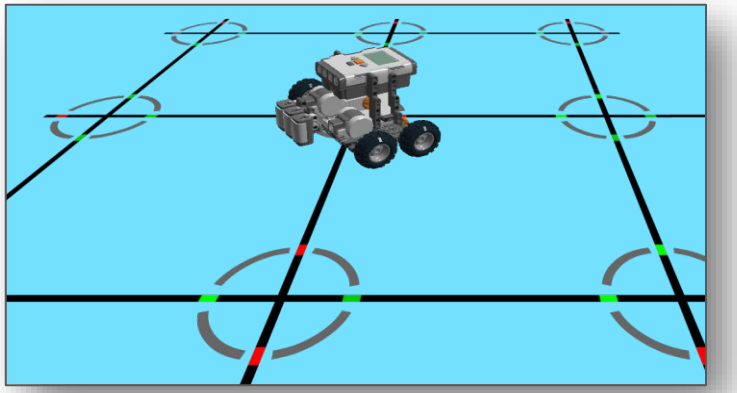


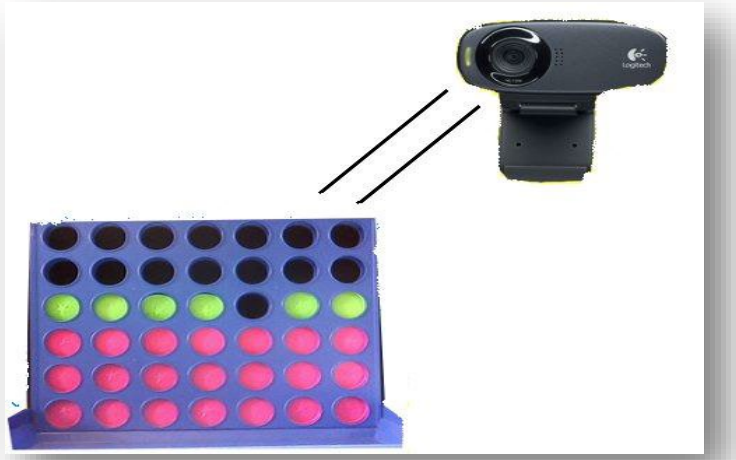
Ollscoil na Gaillimhe
University of Galway

Reinforcement Learning

RL Agent that learns to play UT2004 through trial & error
- Goal: human-like performance

Robots that learn to navigate mazes & solve puzzles





The Future …

Algorithms for learning from mixed media
Systems that can automatically adapt to changing circumstances (adaptive, 'self-healing')
    Software and embedded in hardware
Search engines capable of resolving ambiguity and synthesising results from multiple sources
    What will the weather be like in Finland next week?
    Wolfram|Alpha, IBM's Watson
Cumulative learning and transfer of skills
Active experimentation
Databases and programming languages with built-in learning; Cloud APIs
Sensors everywhere; small & wearable computing

What Do Experts Say About the Future of Machine Learning (and Python)?
https://www.stxnext.com/blog/future-of-machine-learning-and-python-expert-opinions

Ollscoil na Gaillimhe
University of Galway

School of Computer Science

Learning Objectives: Review

If you have been paying close attention, you will now be able to …

1. Discuss definitions of Machine Learning
2. Describe what major categories of ML task entail: classification, regression, clustering, relationship discovery and reinforcement learning
3. Discuss the relationship with Data Mining
4. Explain the Data Mining process
5. Consider current and future applications of Machine Learning and Data Mining

Some of my ML work at NUI Galway:

"*Towards Inherently Adaptive First Person Shooter Agents using Reinforcement Learning*"

"A One-Sided Classification Toolkit with Applications in the Analysis of Spectroscopy Data"

"*Balancing the Performance of a FightingICE Agent Using Reinforcement Learning and Skilled Experience Catalogue*"

"Facial Expression Recognition of Animated Characters using Deep Learning"

"*Robust Classification of High-Dimensional Spectroscopy Data Using Deep Learning and Data Synthesis*"

"*A Multi-class Approach–Building a Visual Classifier based on Textual Descriptions using Zero-Shot Learning*"

Ollscoil na Gaillimhe
University of Galway

School of Computer Science