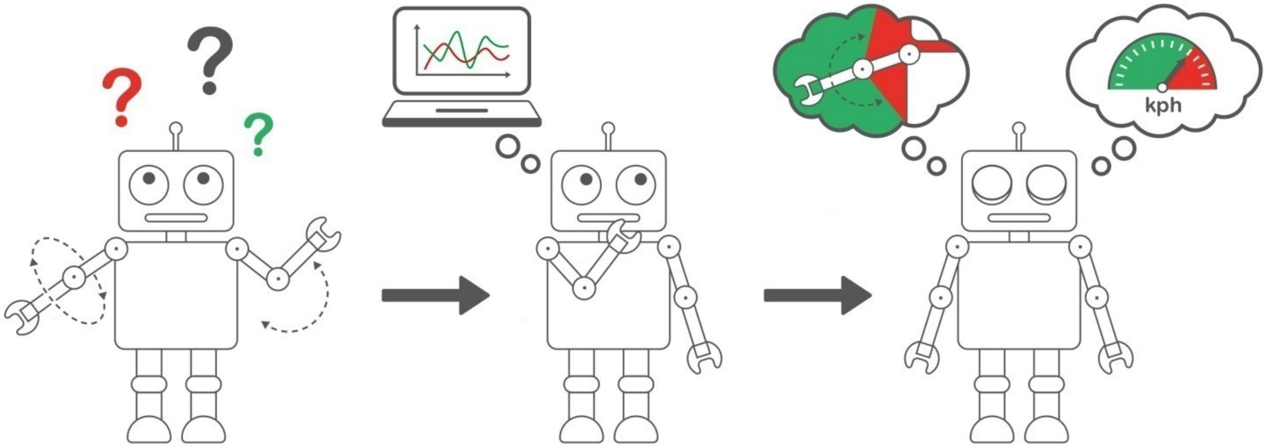


Course on Cyber Security and Deep Learning (July 15th, 2020)

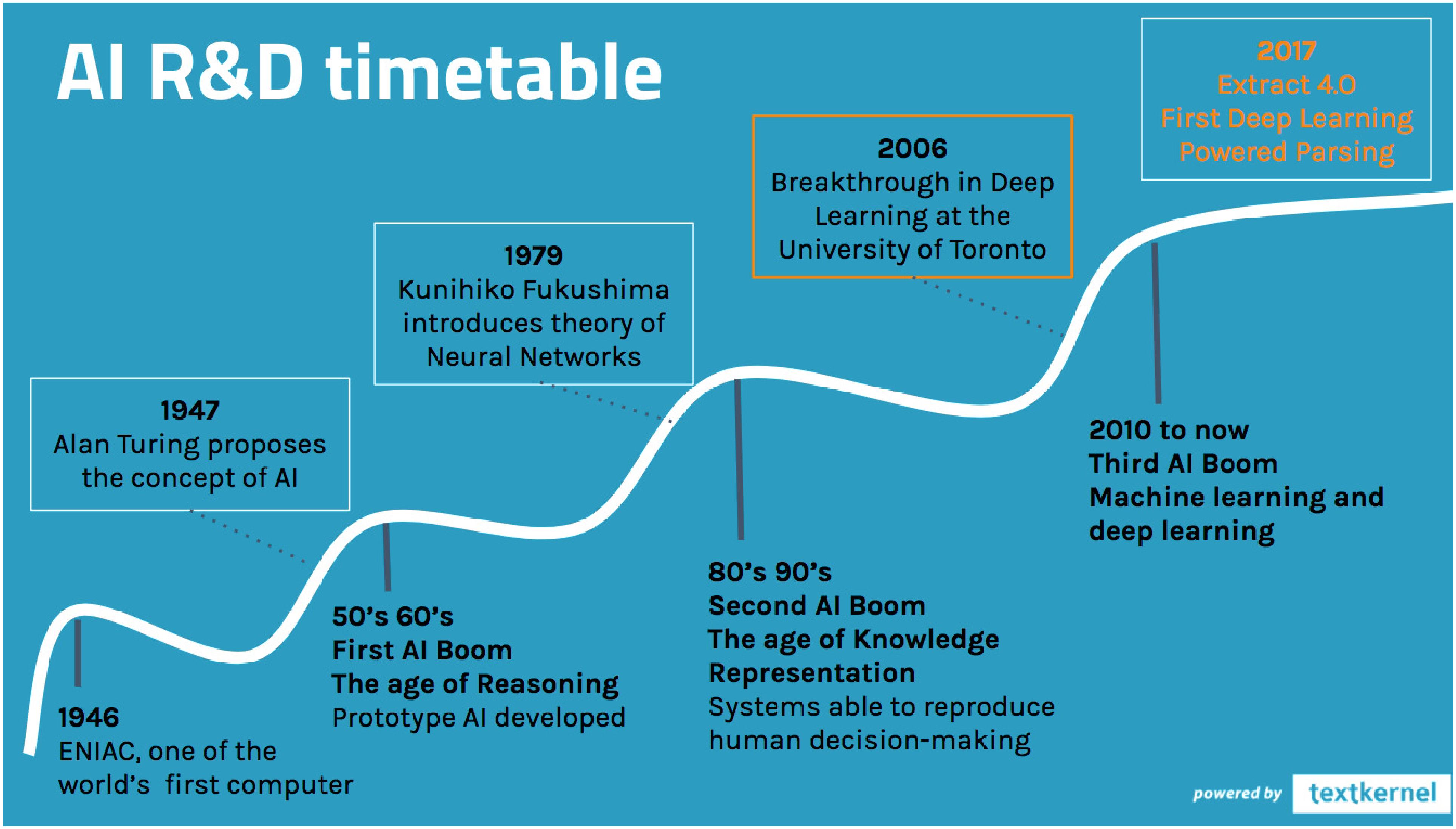
Deep Learning –An Introduction

Dr. Anand Kumar M
Assistant Professor-I,
Department of Information Technology
National Institute of Technology Karnataka
Surathkal

Outline & Content

- AI/ML/DL
- Machine Leanring
- Deep learning?
- Why Deep Learning
- Applications
- Conclusion

AI R&D timetable



1946
ENIAC, one of the
world's first computer

1947
Alan Turing proposes
the concept of AI

50's 60's
First AI Boom
The age of Reasoning
Prototype AI developed

1979
Kunihiko Fukushima
introduces theory of
Neural Networks

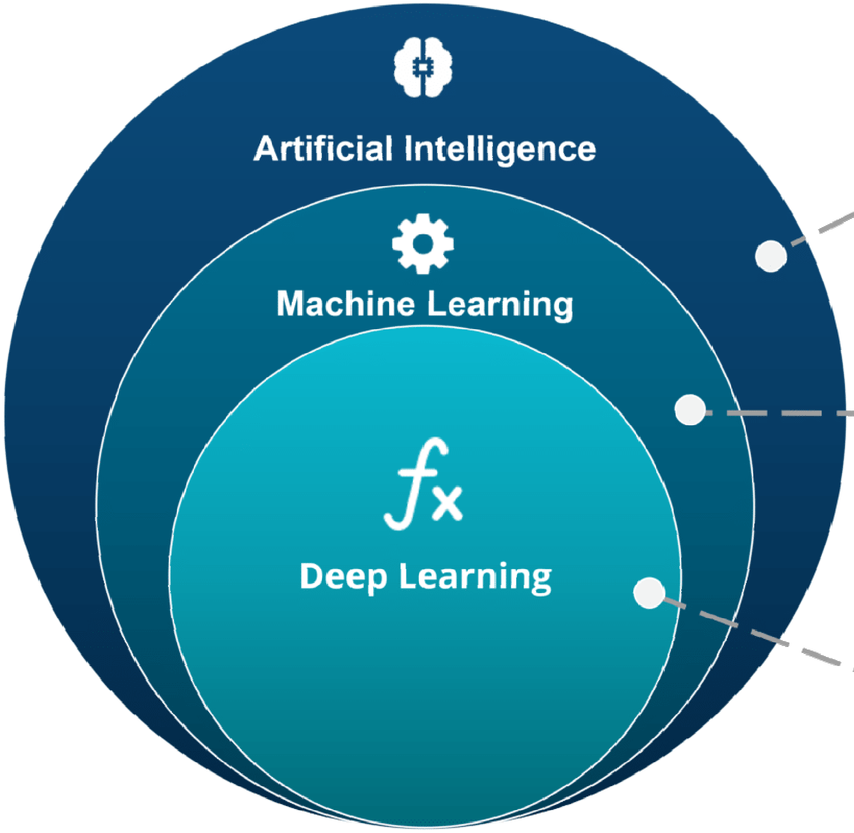
80's 90's
Second AI Boom
The age of Knowledge Representation
Systems able to reproduce
human decision-making

2006
Breakthrough in Deep
Learning at the
University of Toronto

2010 to now
Third AI Boom
Machine learning and deep learning

2017
Extract 4.0
First Deep Learning
Powered Parsing

powered by textkernel



ARTIFICIAL INTELLIGEN

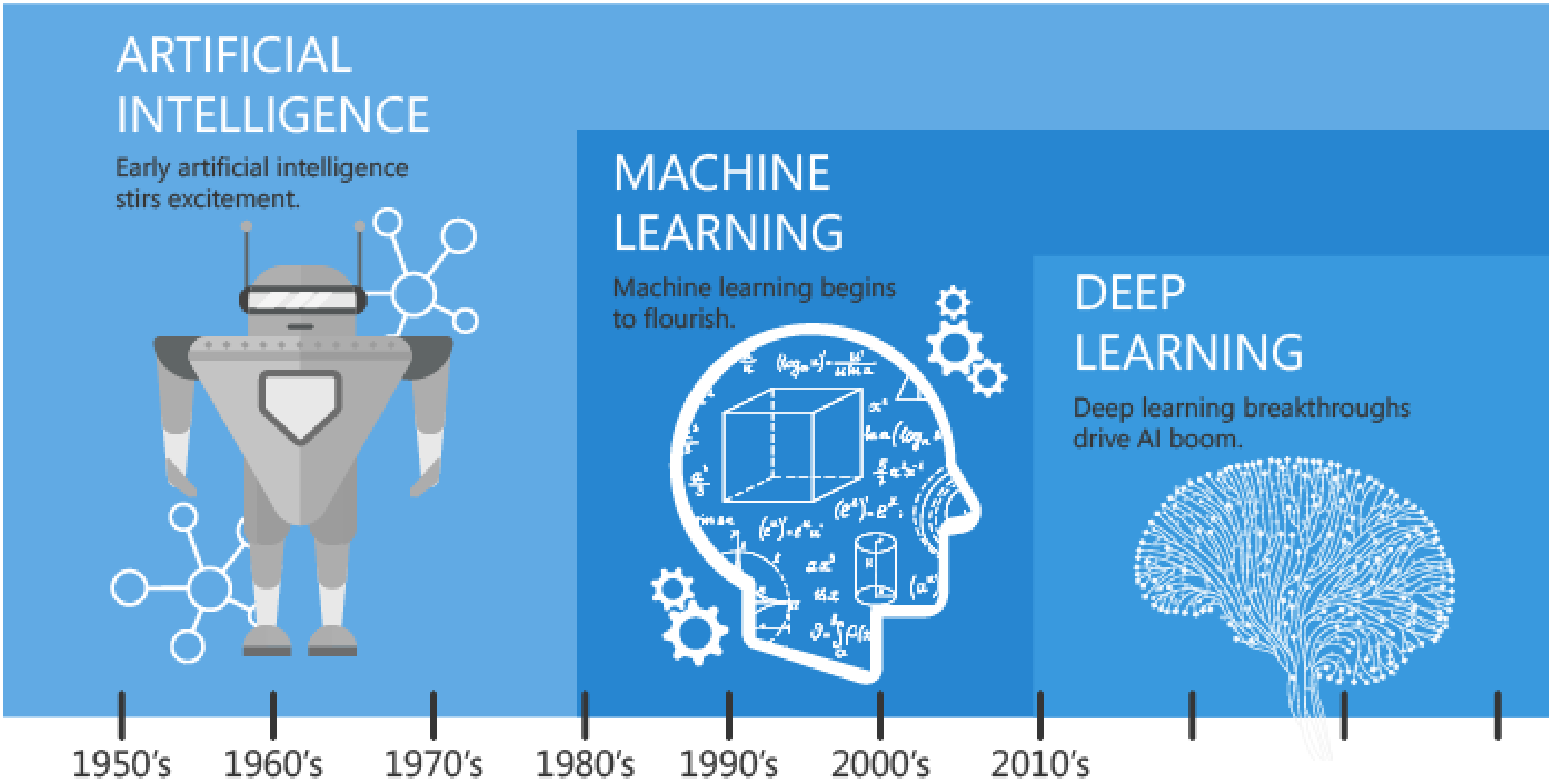
A technique which enables mach
to mimic human behaviour

MACHINE LEARNING

Subset of AI technique which us
statistical methods to enable ma
to improve with experience

DEEP LEARNING

Subset of ML which make the
computation of multi-layer neura
network feasible



ARTIFICIAL INTELLIGENCE

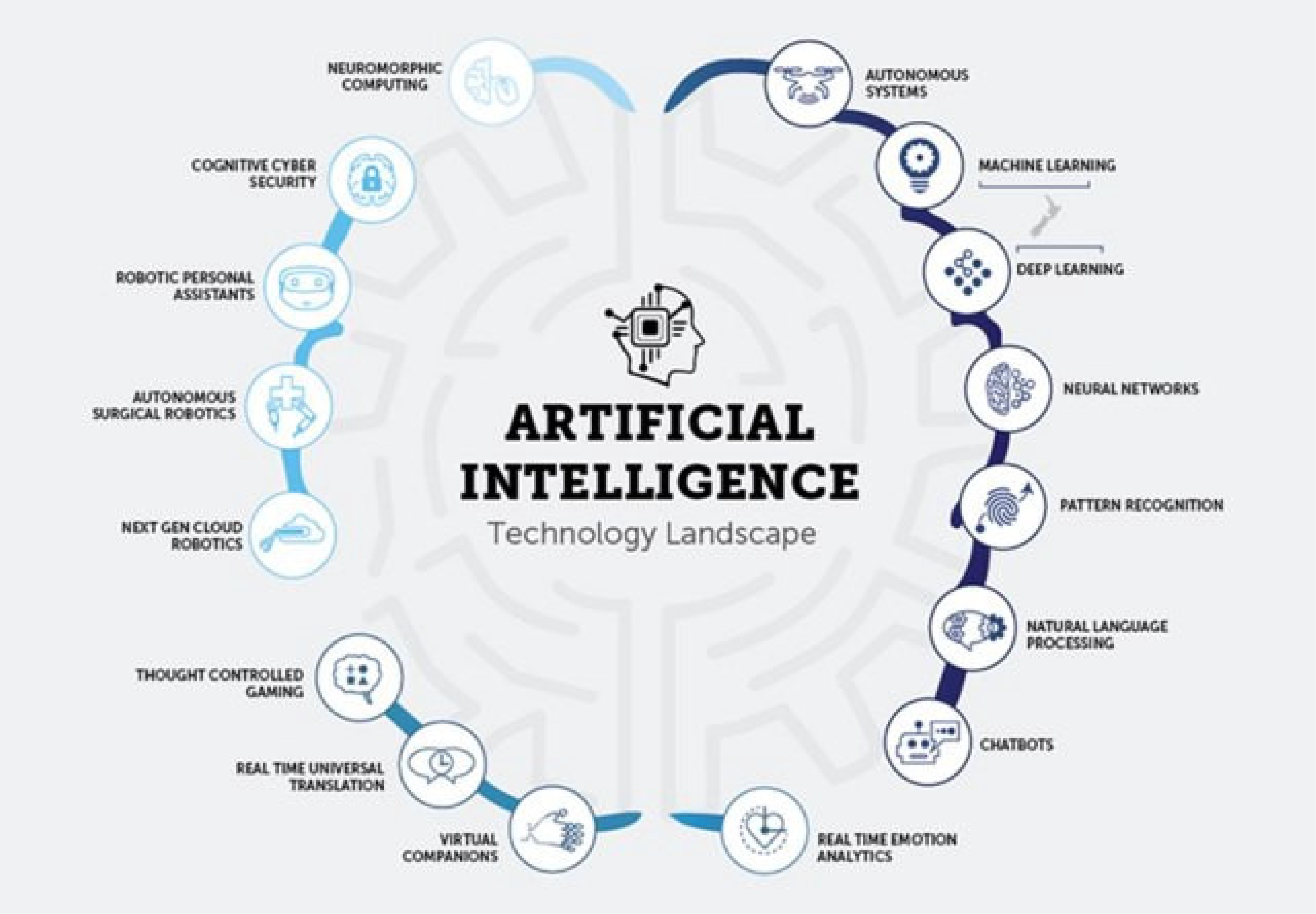
Early artificial intelligence stirs excitement.

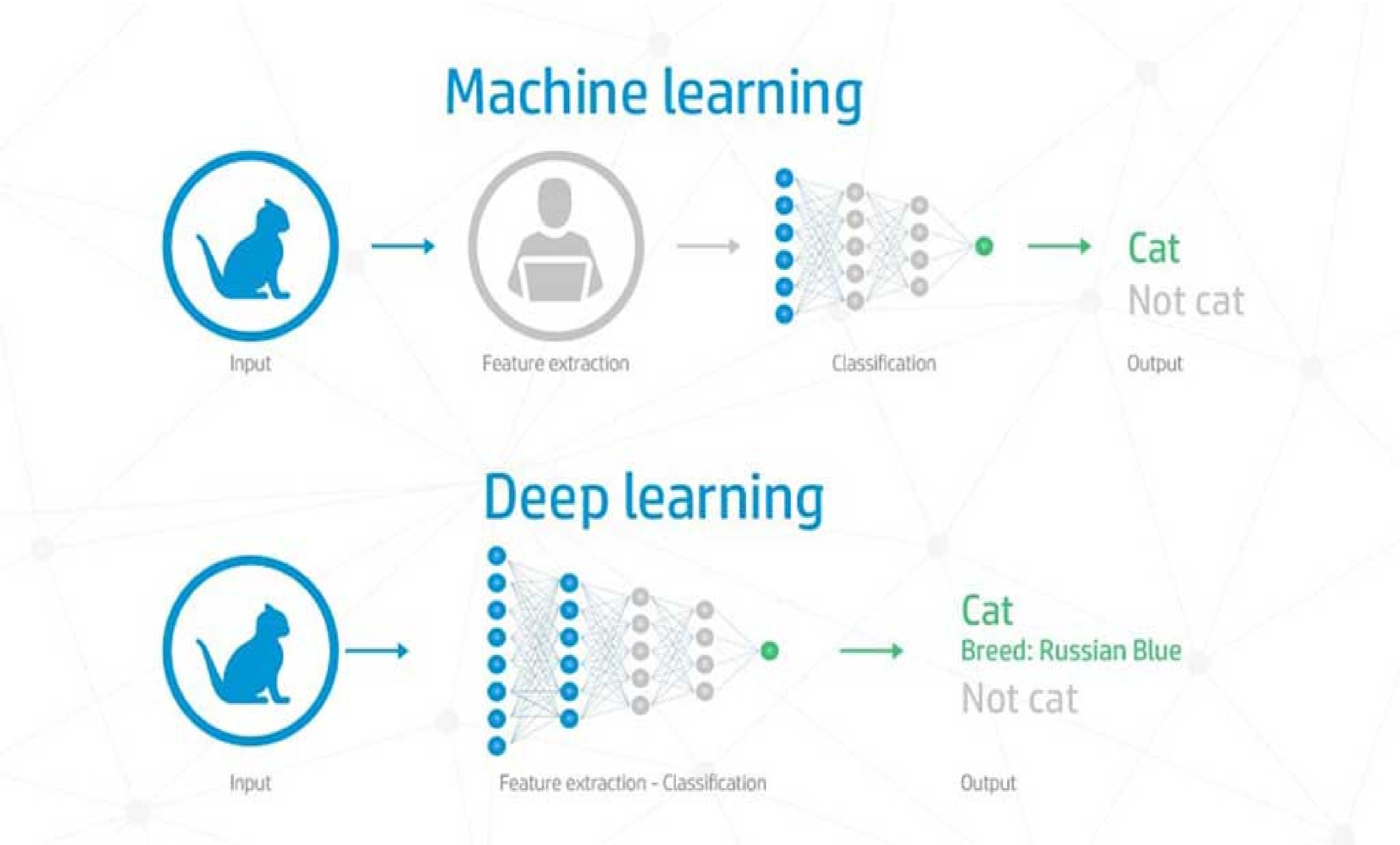
MACHINE LEARNING

Machine learning begins to flourish.

DEEP LEARNING

Deep learning breakthroughs drive AI boom.

ARTIFICIAL INTELLIGENCE
Technology Landscape
NEUROMORPHIC COMPUTING
COGNITIVE CYBER SECURITY
ROBOTIC PERSONAL ASSISTANTS
AUTONOMOUS SURGICAL ROBOTICS
NEXT GEN CLOUD ROBOTICS
THOUGHT CONTROLLED GAMING
REAL TIME UNIVERSAL TRANSLATION
VIRTUAL COMPANIONS
AUTONOMOUS SYSTEMS
MACHINE LEARNING
DEEP LEARNING
NEURAL NETWORKS
PATTERN RECOGNITION
NATURAL LANGUAGE PROCESSING
CHATBOTS
REAL TIME EMOTION ANALYTICS



Machine learning

Input

Feature extraction

Classification

Cat

Not cat

Output

Deep learning

Input

Feature extraction - Classification

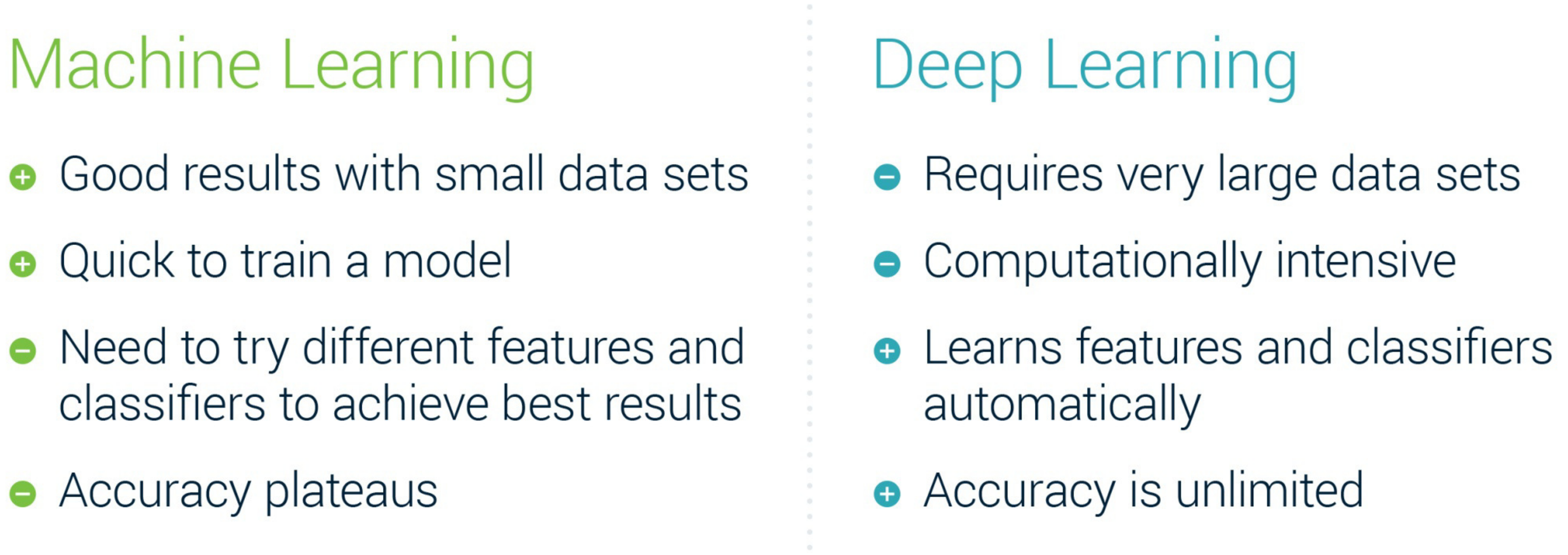
Cat

Breed: Russian Blue

Not cat

Output

ML vs DL



Machine Learning

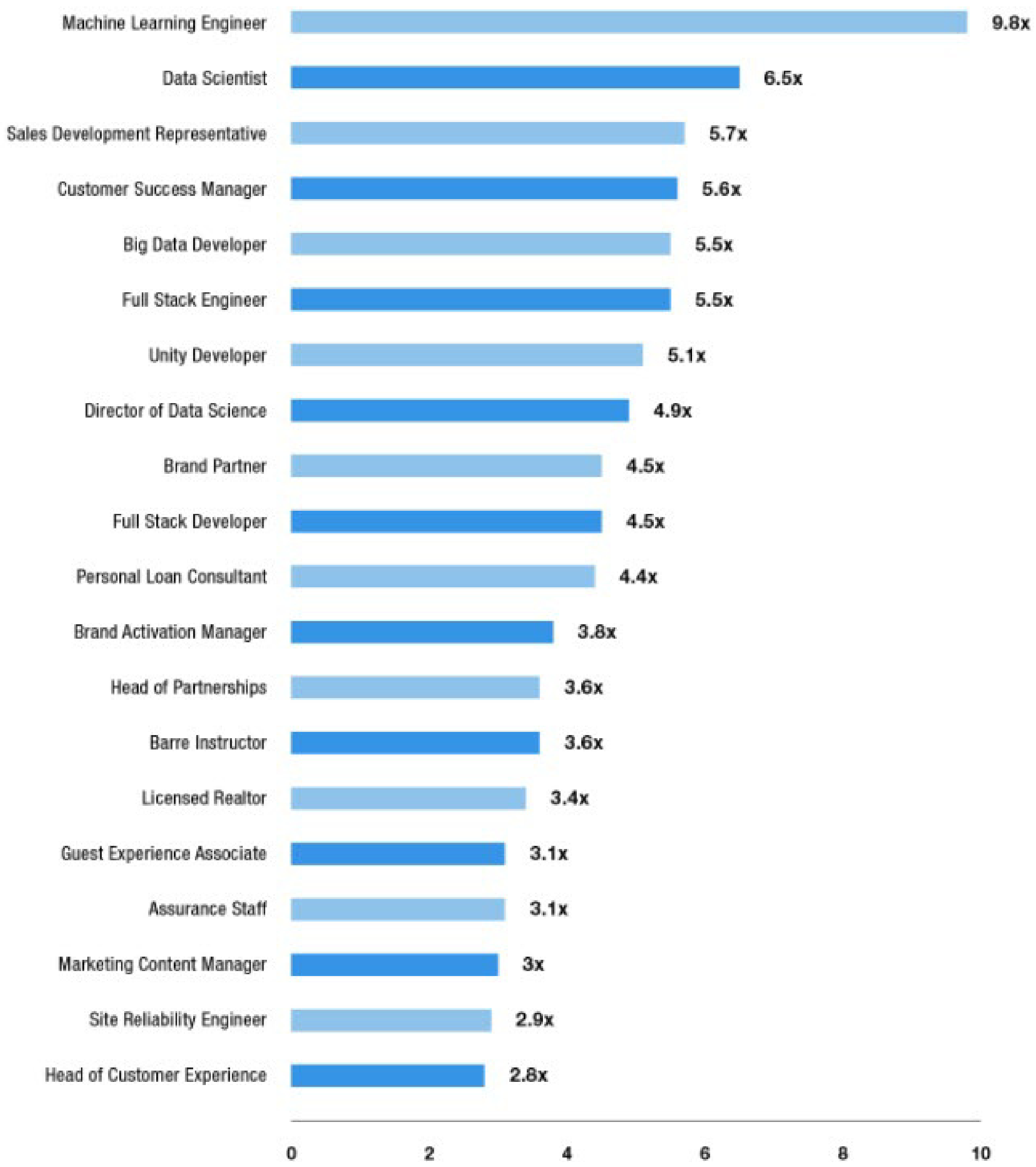
- ⊕ Good results with small data sets
- ⊕ Quick to train a model
- ⊖ Need to try different features and classifiers to achieve best results
- ⊖ Accuracy plateaus

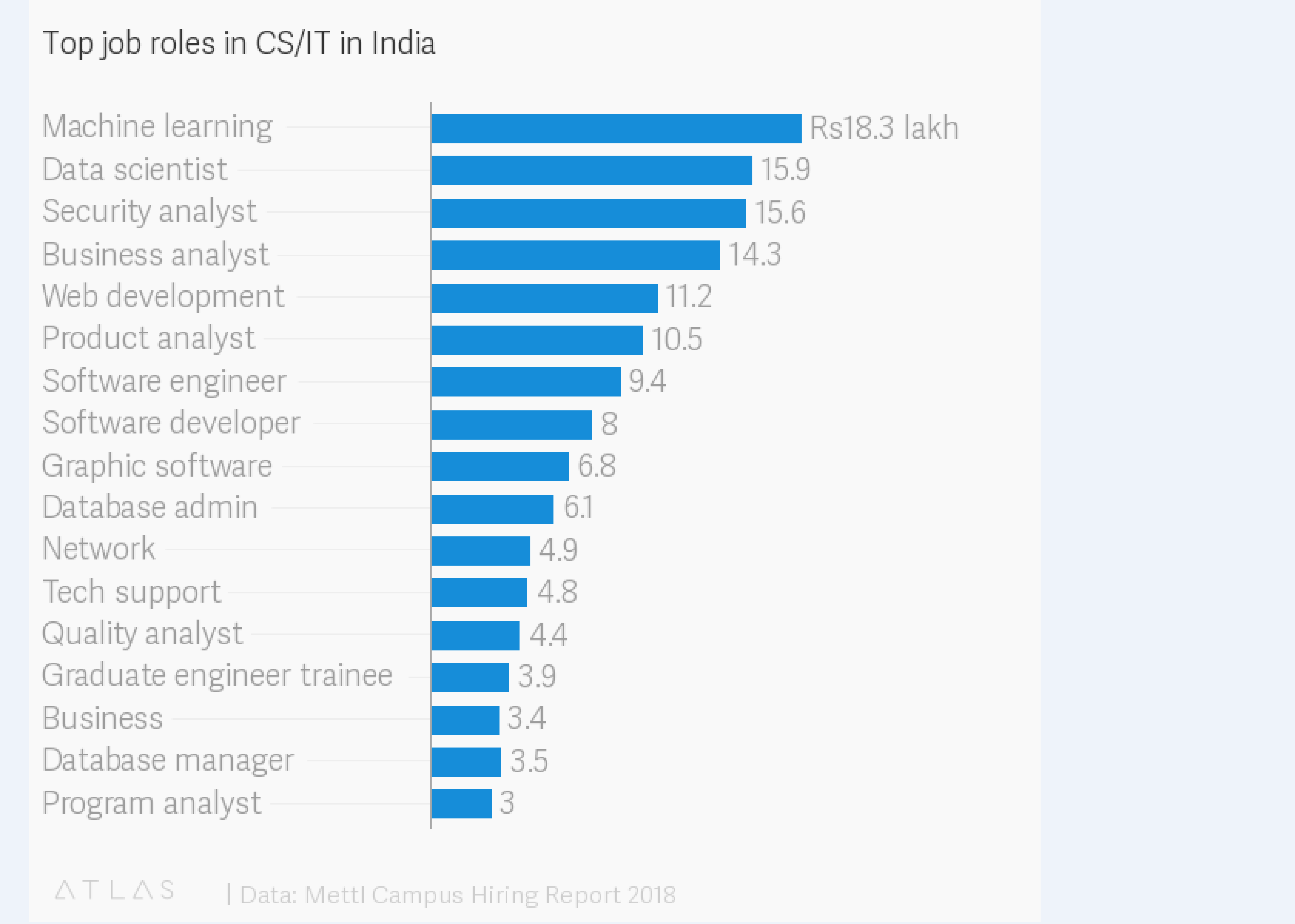
Deep Learning

- ⊖ Requires very large data sets
- ⊖ Computationally intensive
- ⊕ Learns features and classifiers automatically
- ⊕ Accuracy is unlimited

Top 20 Emerging Jobs

LinkedIn Economic Graph




Top job roles in CS/IT in India

| Data: Mettl Campus Hiring Report 2018

WORLD
ECONOMIC
FORUM

COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The Jobs Landscape in 2022

emerging roles, global change by 2022

133 Million

Top 10 Emerging

1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. General and Operations Managers
4. Software and Applications Developers and Analysts
5. Sales and Marketing Professionals
6. Big Data Specialists
7. Digital Transformation Specialists
8. New Technology Specialists
9. Organisational Development Specialists
10. Information Technology Services

declining roles, global change by 2022

75 Million

Top 10 Declining

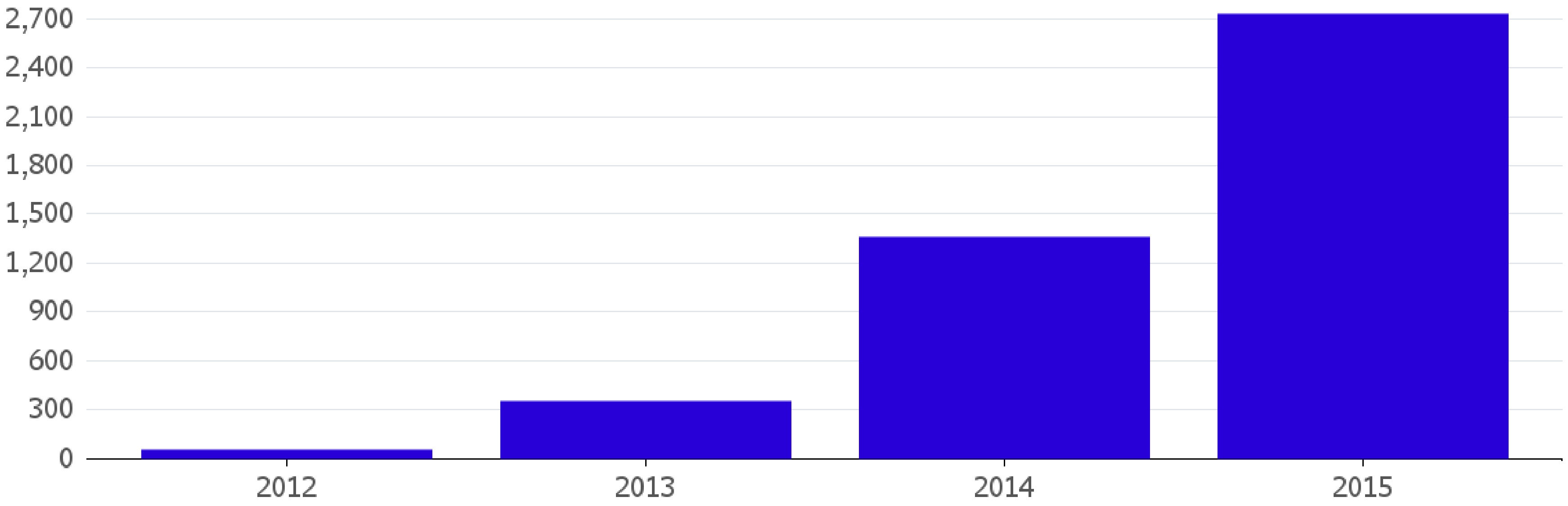
1. Data Entry Clerks
2. Accounting, Bookkeeping and Payroll Clerks
3. Administrative and Executive Secretaries
4. Assembly and Factory Workers
5. Client Information and Customer Service Workers
6. Business Services and Administration Managers
7. Accountants and Auditors
8. Material-Recording and Stock-Keeping Clerks
9. General and Operations Managers
10. Postal Service Clerks

Source: Future of Jobs Report 2018, World Economic Forum

Deep Learning at Google

Artificial Intelligence Takes Off at Google

Number of software projects within Google that uses a key AI technology, called Deep Learning.

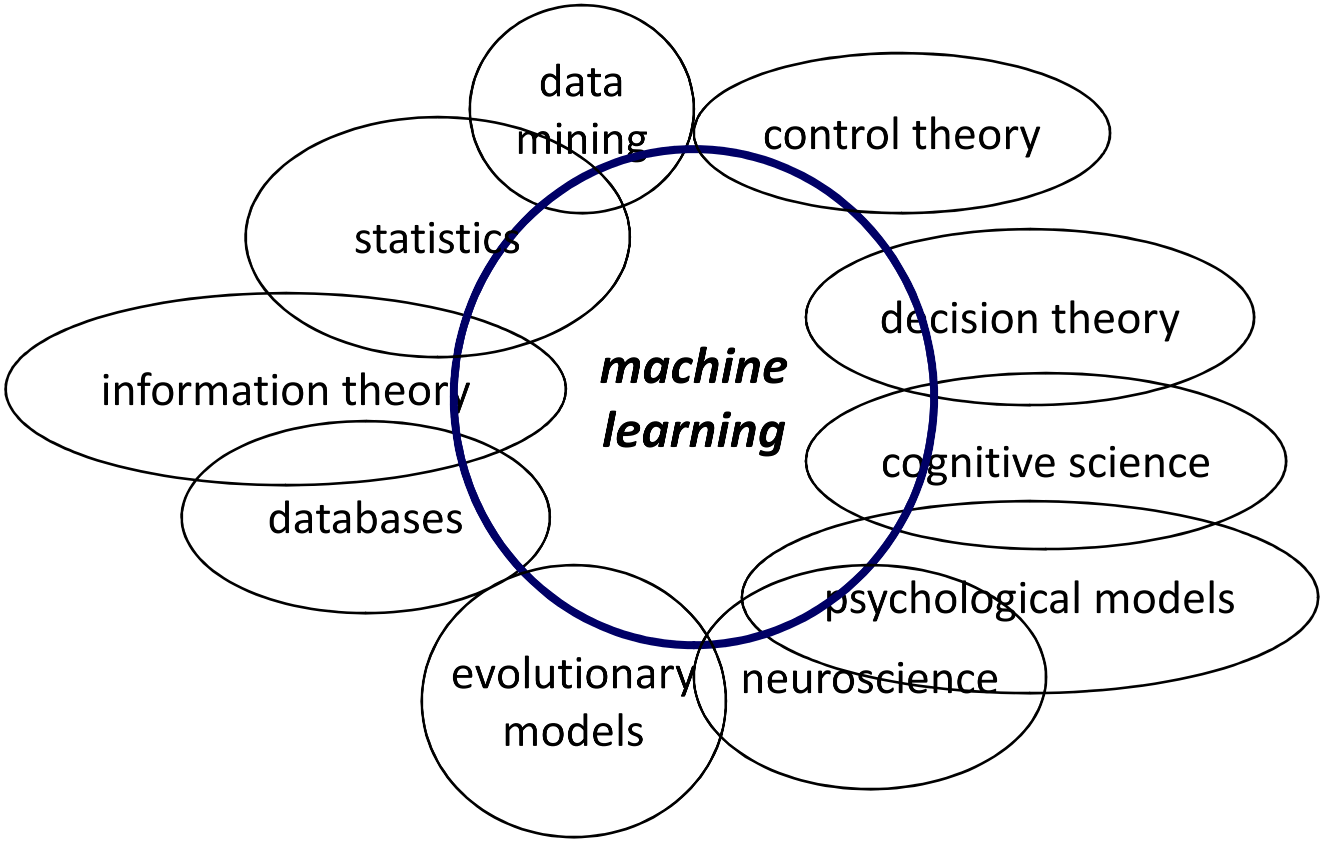


Source: Google

Note: 2015 data does not incorporate data from Q4

Bloomberg

Related Fields



Machine learning is primarily concerned with the accuracy and effectiveness of the *computer system*.

What is Machine Learning?

- It is very hard to write programs that solve problems like **recognizing a face.**
  - We don't know what program to write because we don't know how our brain does it.
  - Even if we had a good idea about how to do it, the program might be awfully complicated.
- Instead of writing a program by hand, we ***collect lots of examples*** that specify the correct output for a given input.
- A machine learning algorithm then takes these examples and produces a program that does the job.
  - The program produced by the ***learning algorithm may look very different from a typical hand-written program***. It may contain millions of numbers.
  - If we do it right, the ***program works for new cases as well as the ones we trained it on.***

Machine Learning

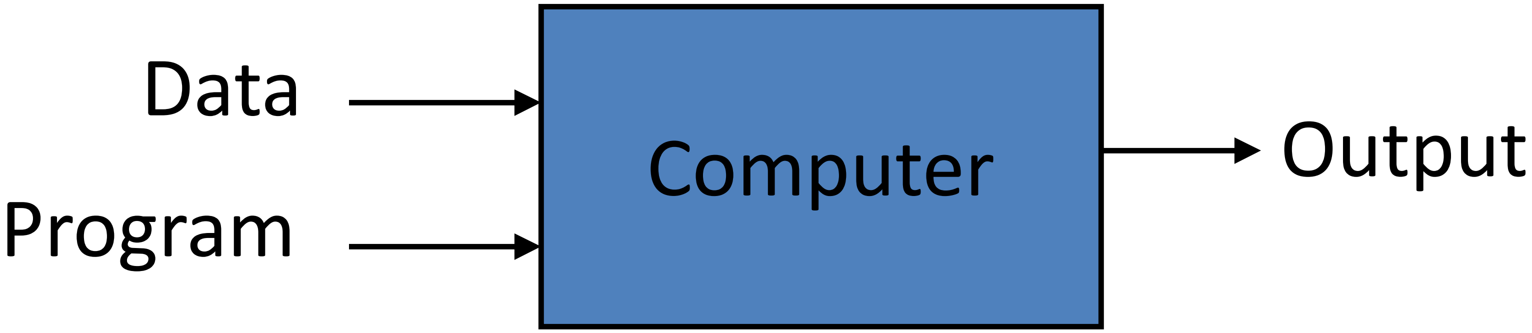
- **Herbert Alexander Simon**: "Learning is any process by which a system improves performance from experience."
- "Machine Learning is concerned with computer programs that automatically improve their performance through experience. "

Herbert Simon
Turing Award 1975
Nobel Prize in Economics 1978

Why now?

- Flood of available data (especially with the advent of the Internet)
- Increasing computational power
- Growing progress in available algorithms and theory developed by researchers
- Increasing support from industries

Traditional Programming



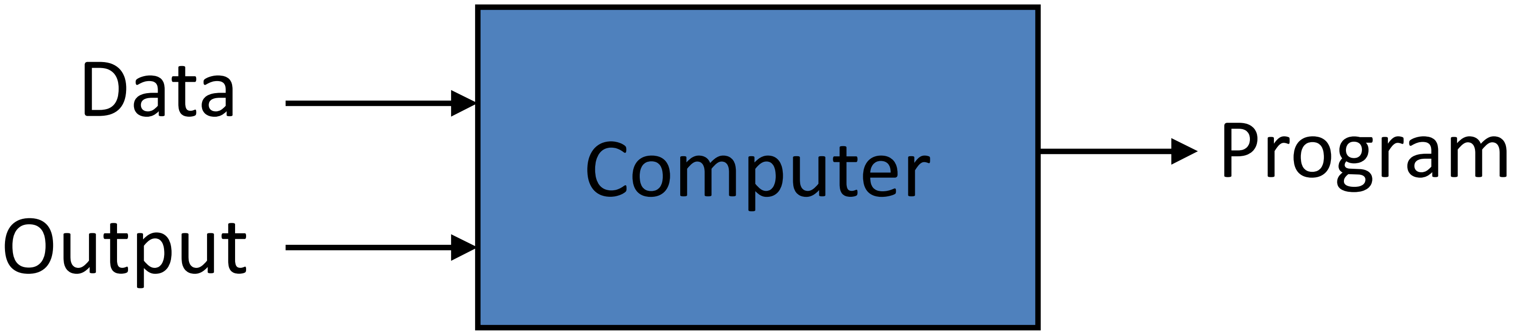
Data

Program

Computer

Output

Machine Learning



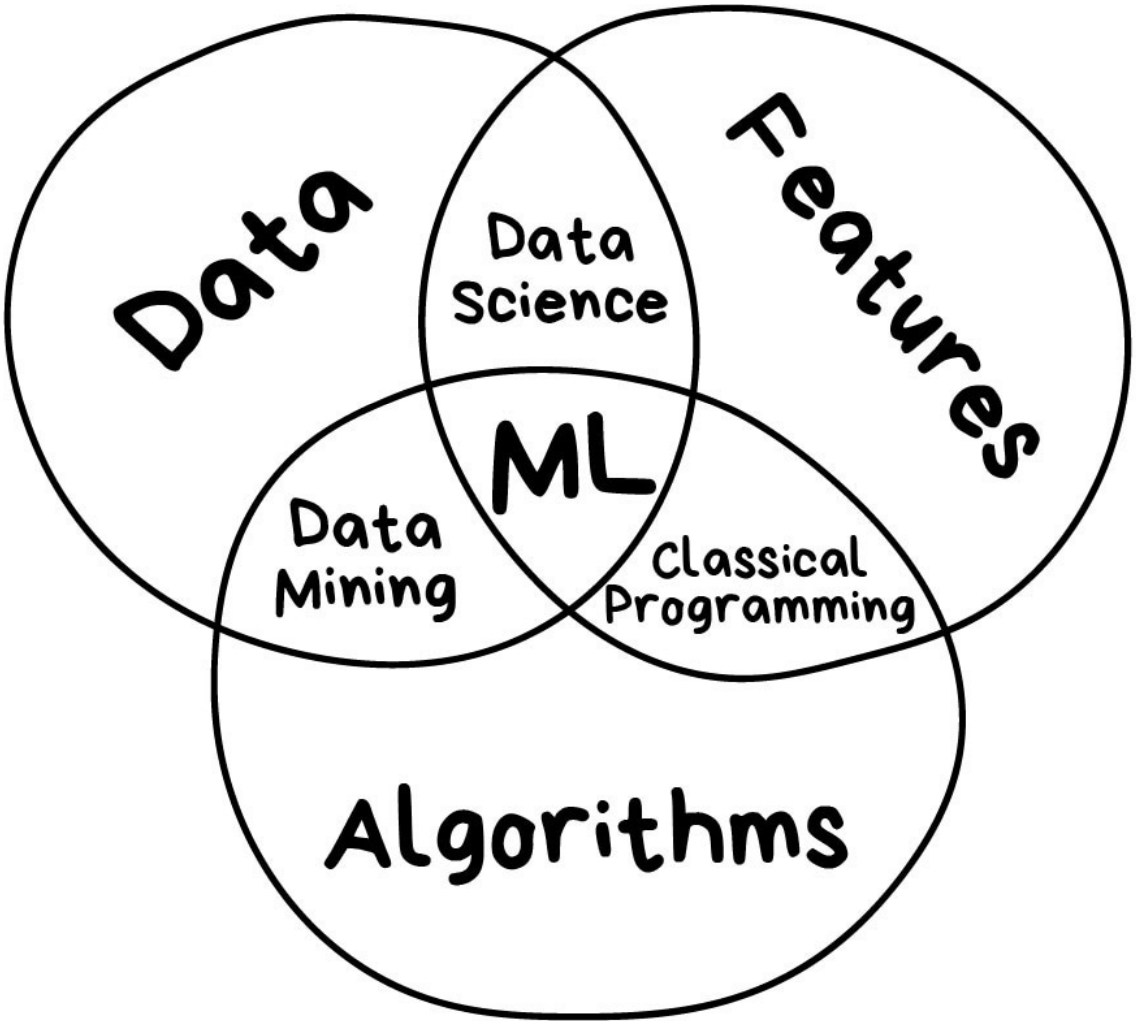
Data

Output

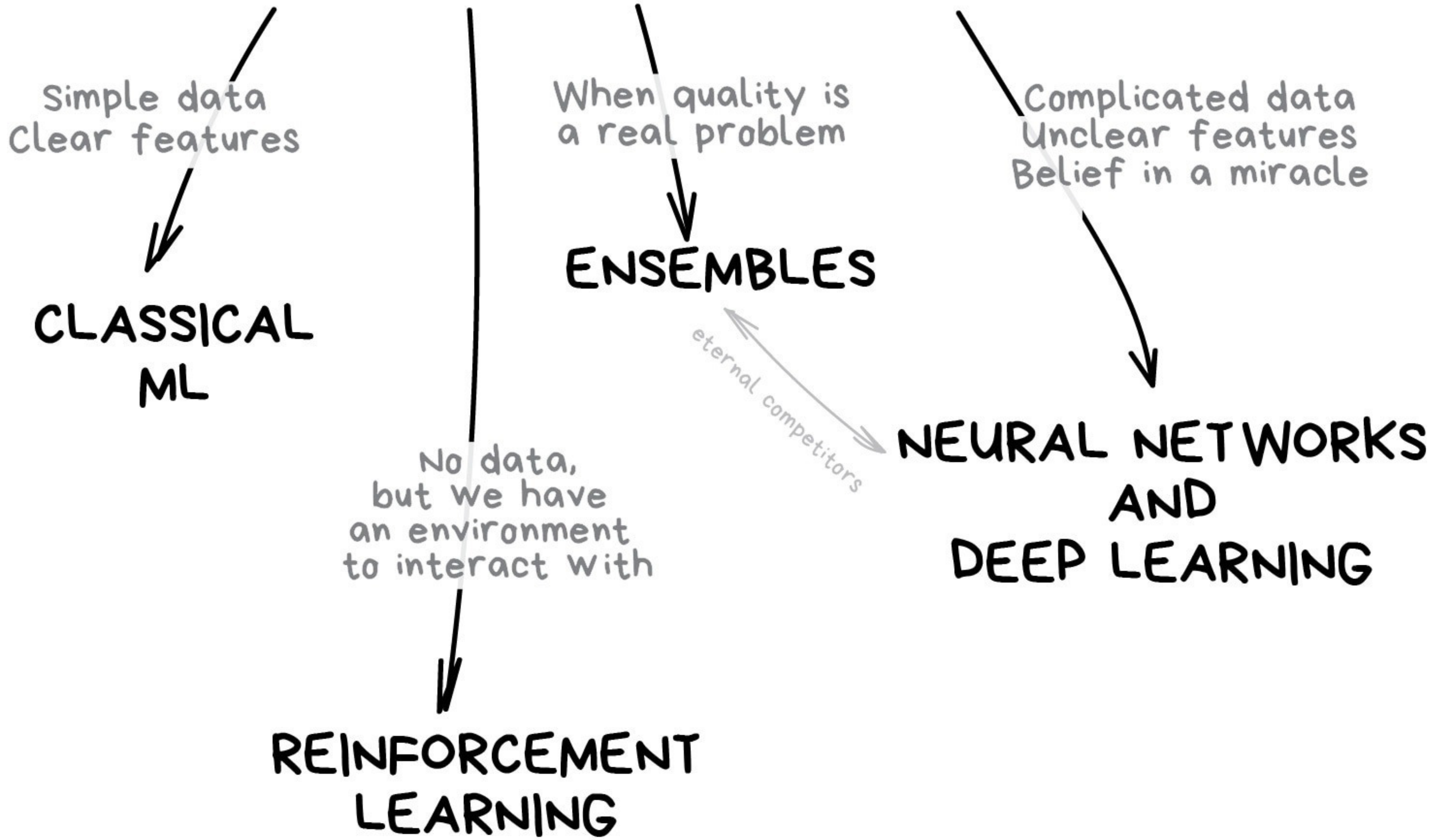
Computer

Program

Three components of machine learning



THE MAIN TYPES OF MACHINE LEARNING



Simple data
Clear features

CLASSICAL ML

No data,
but we have
an environment
to interact with

REINFORCEMENT LEARNING

When quality is
a real problem

ENSEMBLES

eternal competitors

Complicated data
Unclear features
Belief in a miracle

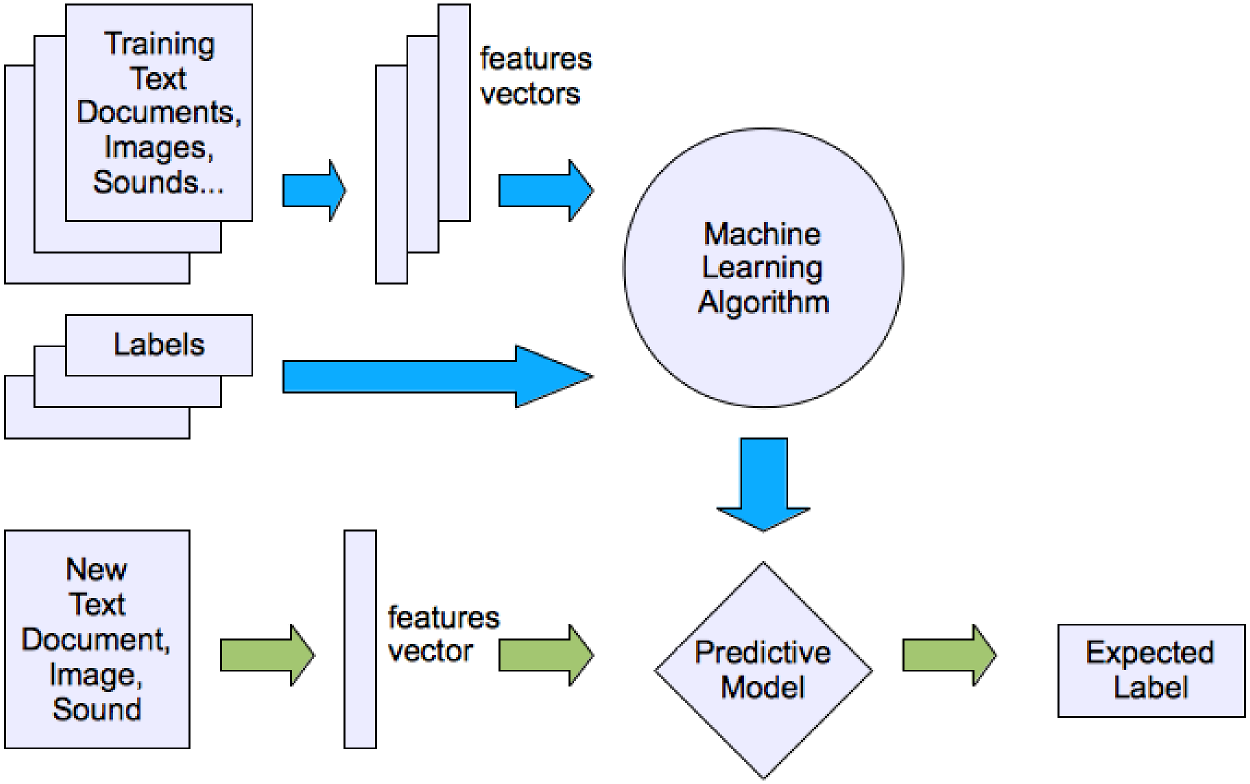
NEURAL NETWORKS AND DEEP LEARNING

Data

- Want to detect spam? Get samples of spam messages. Want to forecast stocks? Find the price history. Want to find out user preferences?

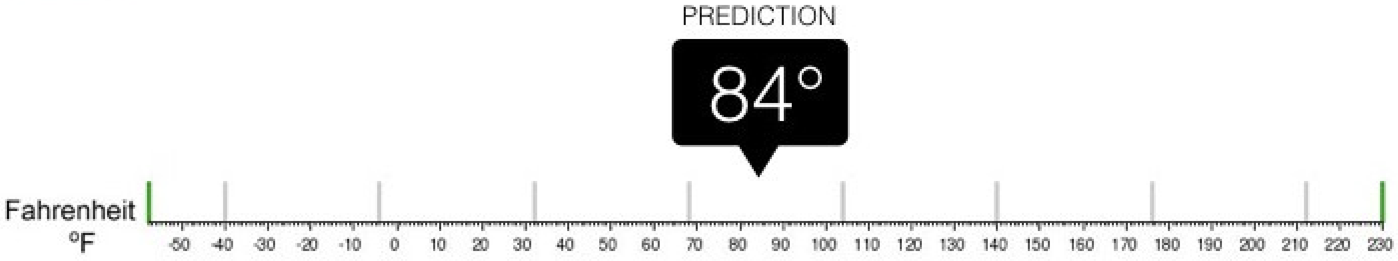
Machine learning structure

- Supervised learning


Training Text Documents, Images, Sounds...
features vectors
Labels
Machine Learning Algorithm
New Text Document, Image, Sound
features vector
Predictive Model
Expected Label

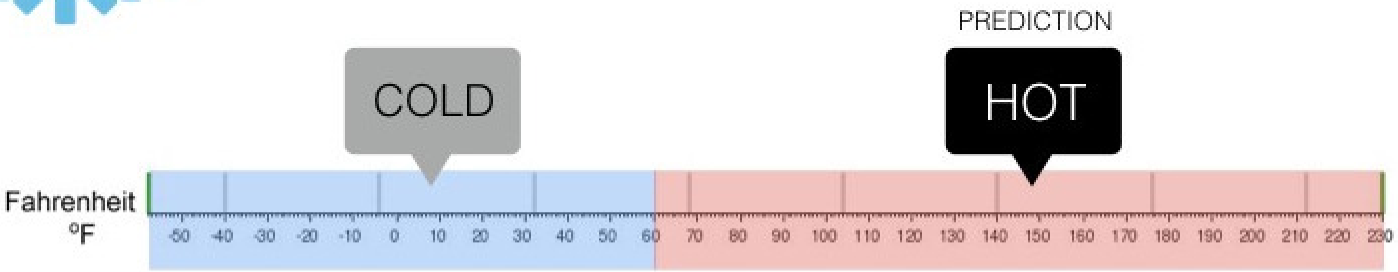
Regression

What is the temperature going to be tomorrow?

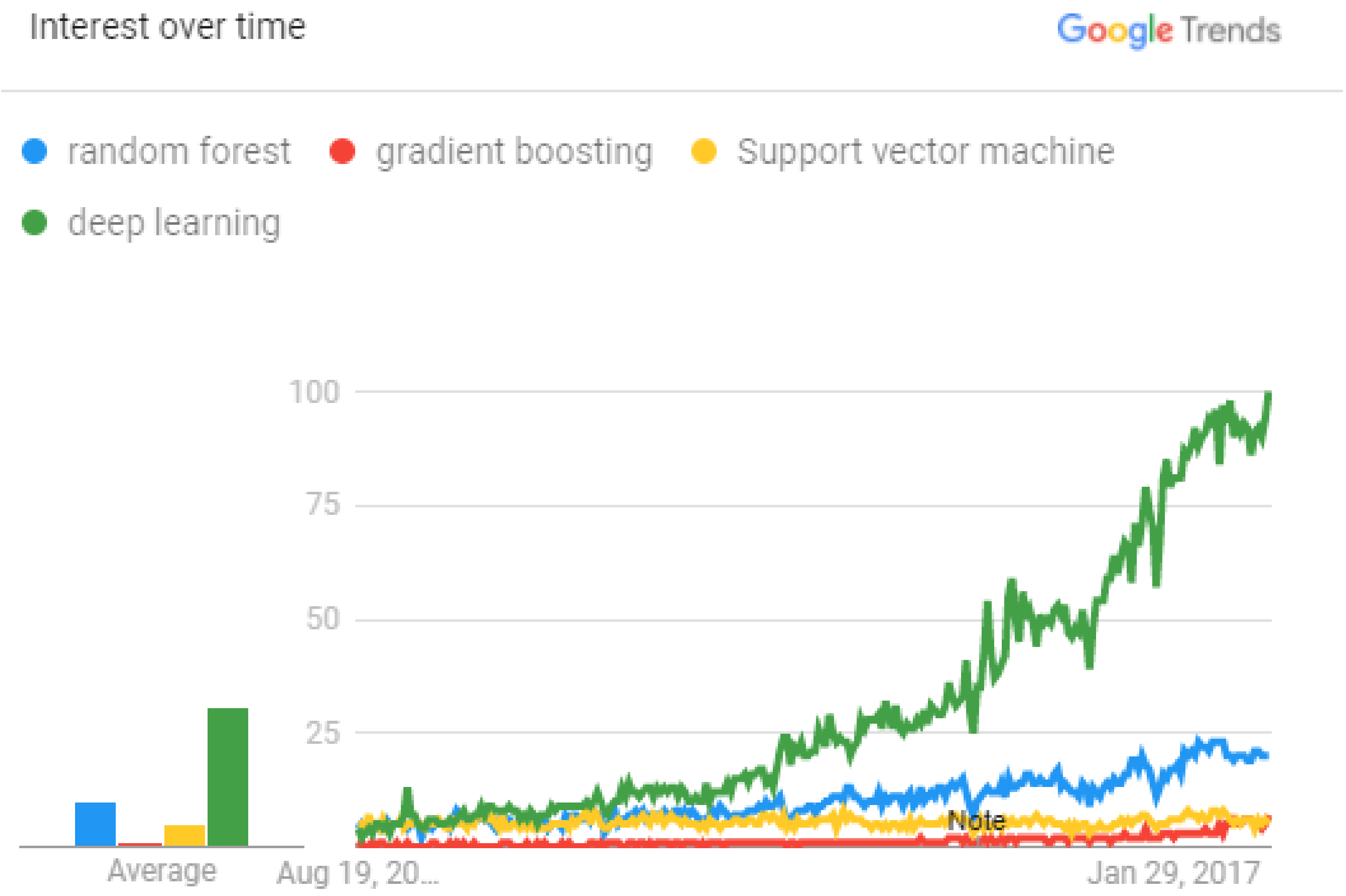


Classification

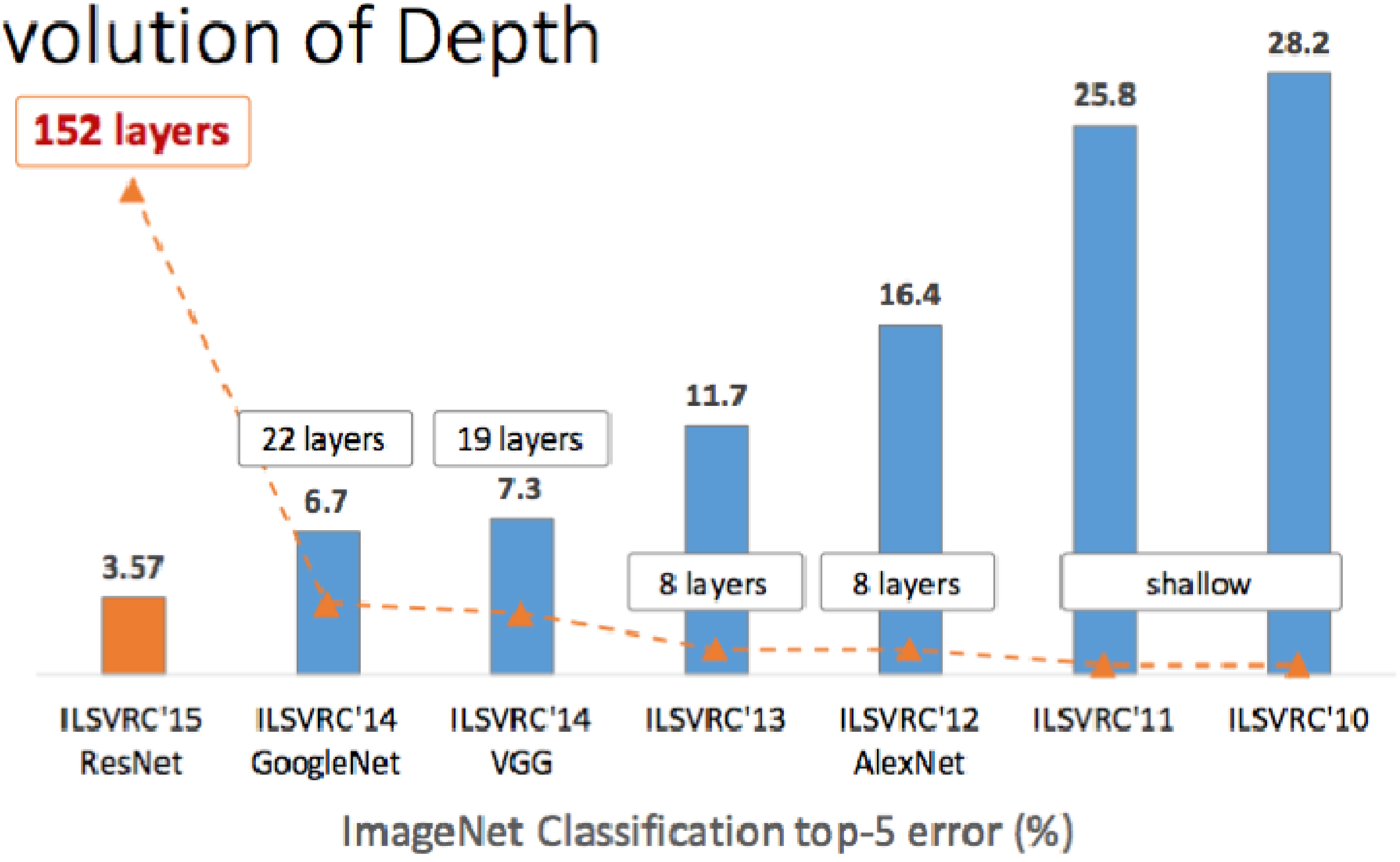
Will it be Cold or Hot tomorrow?



Google trends



Revolution of Depth



So, 1. **what exactly is deep learning** ?

And, 2. **why is it generally better** than other methods on image, speech and certain other types of data?

So, 1. **what exactly is deep learning** ?

And, 2. **why is it generally better** than other methods on image, speech and certain other types of data?

The short answers

1. **‘Deep Learning’ means** using a **neural network** with **several layers of nodes** between input and output

2. **the series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.**

hmmm… OK, but:

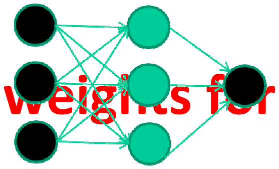
3. **multilayer neural networks have been around for 25 years. What’s actually new?**

hmmm… OK, but:

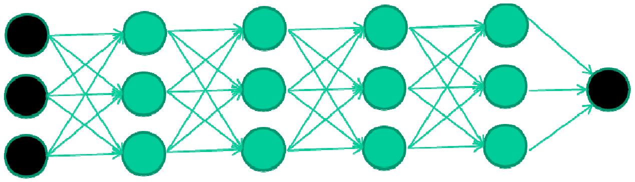
3. **multilayer neural networks have been around for 25 years. What's actually new?**

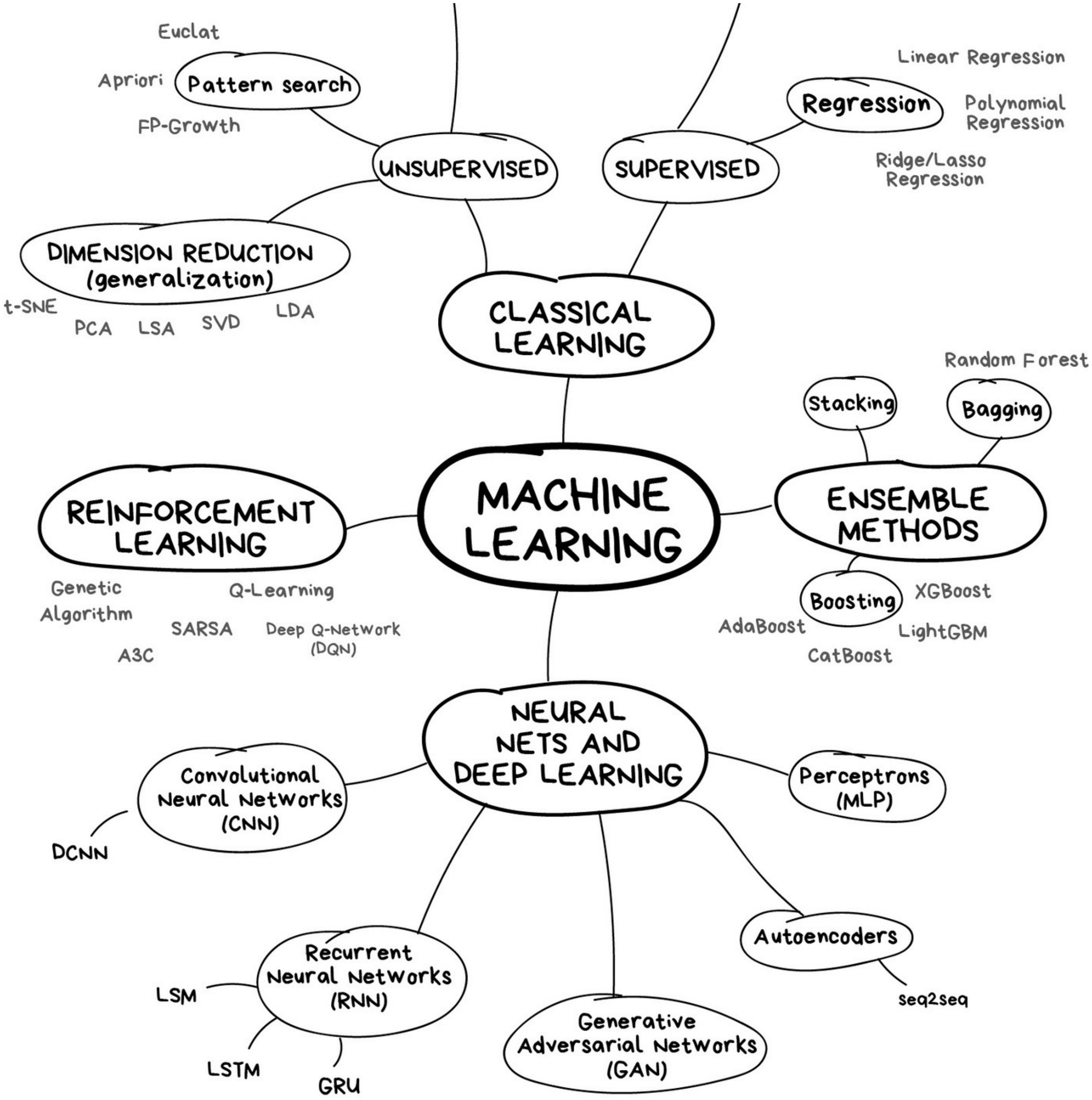
we have always had good algorithms for learning the weights in networks with 1 hidden layer

but these algorithms are not good at learning the weights for networks with more hidden layers

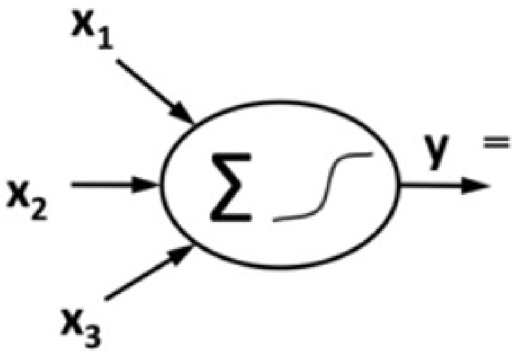


what's new is: **algorithms for training many-later networks**




MACHINE LEARNING
CLASSICAL LEARNING
UNSUPERVISED
Pattern search
Euclat
Apriori
FP-Growth
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
SUPERVISED
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
ENSEMBLE METHODS
Stacking
Bagging
Random Forest
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
REINFORCEMENT LEARNING
Genetic Algorithm
Q-Learning
SARSA
Deep Q-Network (DQN)
A3C
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq

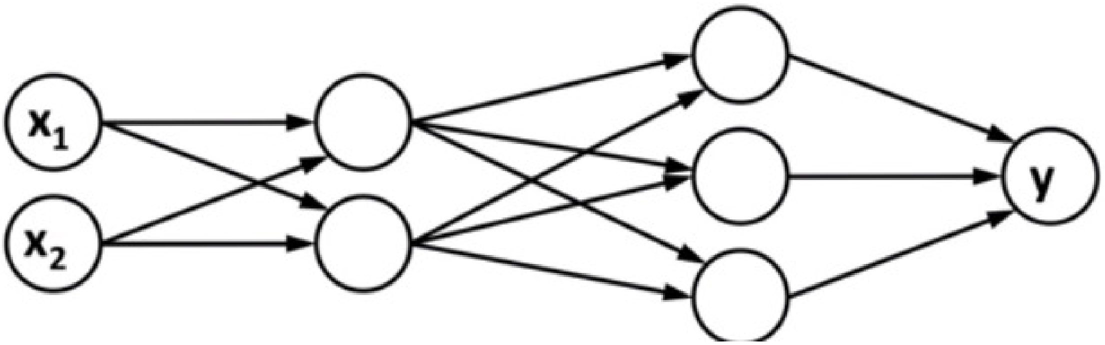
Types of Neural Networks



$$= \sigma\left(b + \sum_i w_{ij} x_i\right)$$

Single neuron: perceptron,

linear / logistic regression

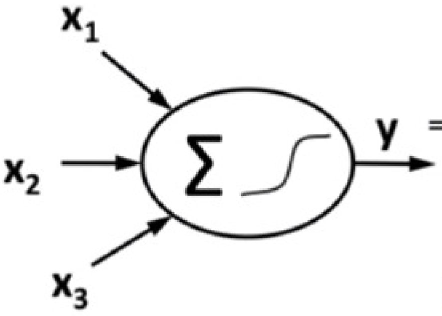


Feed-forward netwo

(no cycles) -- non-lin

classification & regre

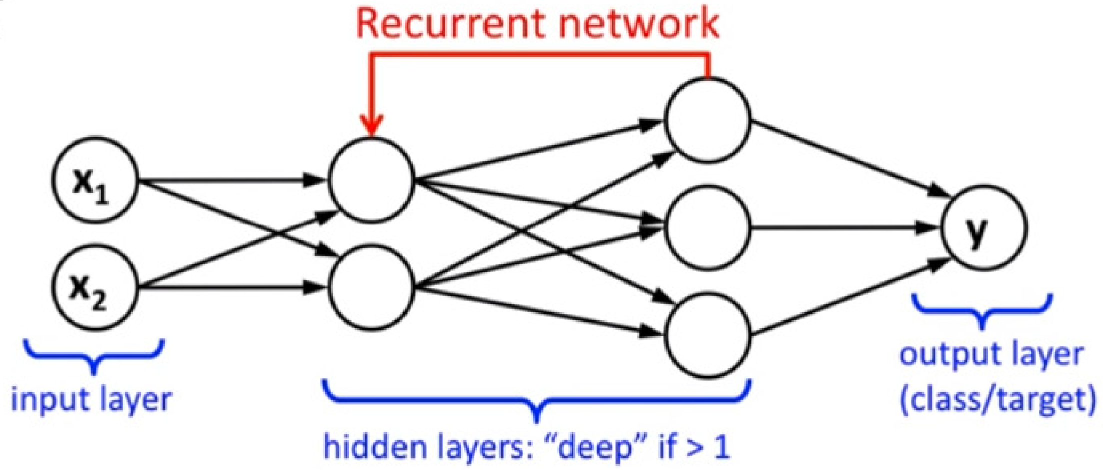
Types of Neural Networks



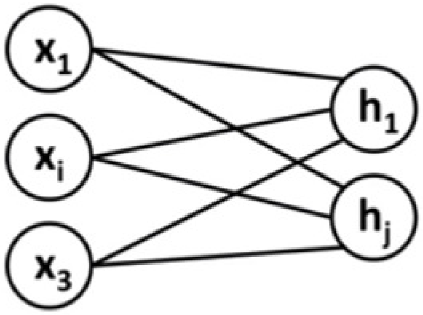
$$y = \sigma\left(b + \sum_i w_{ij} x_i\right)$$

Single neuron: perceptron,

linear / logistic regression

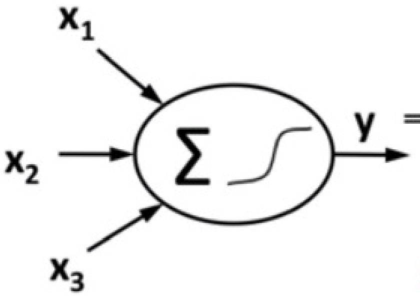


Feed-forward network
(no cycles) -- non-linear classification & regression



Symmetric (RBM)
unsupervised, trained to maximize likelihood of input data

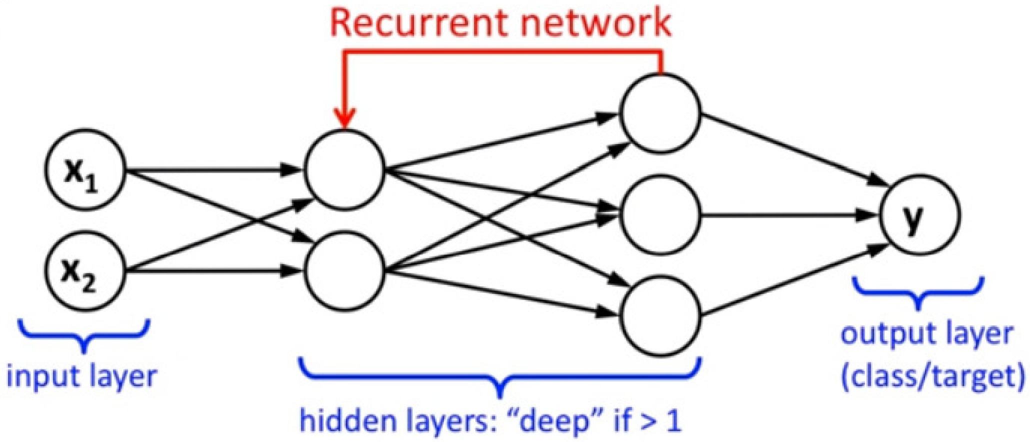
Types of Neural Networks



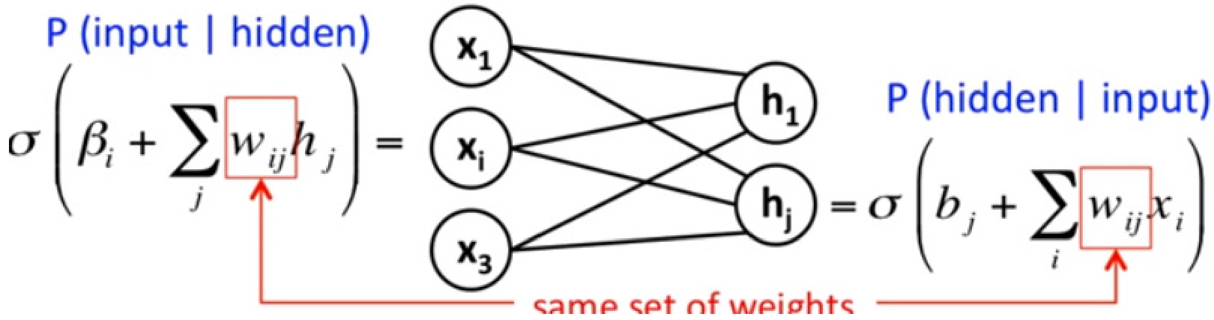
$$y = \sigma\left(b + \sum_i w_{ij} x_i\right)$$

Single neuron: perceptron,

linear / logistic regression



Feed-forward network
(no cycles) -- non-linear classification & regression

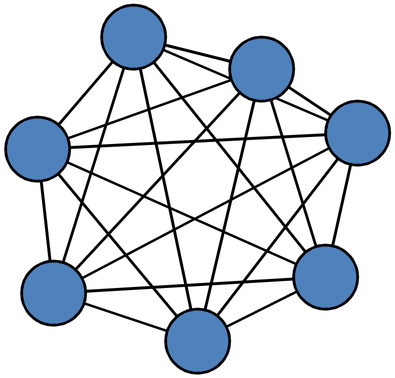


$$\sigma\left(\beta_i + \sum_j \boxed{w_{ij}} h_j\right) =$$

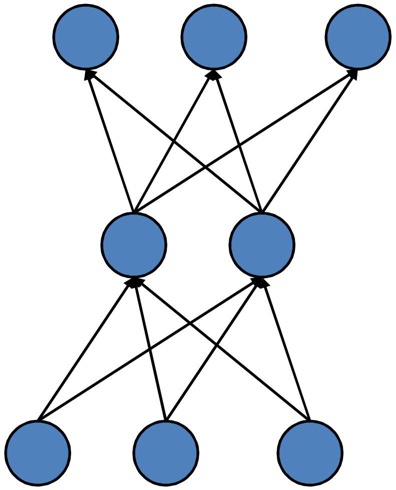
$$= \sigma\left(b_j + \sum_i \boxed{w_{ij}} x_i\right)$$

Symmetric (RBM)
unsupervised, trained to maximize likelihood of input data

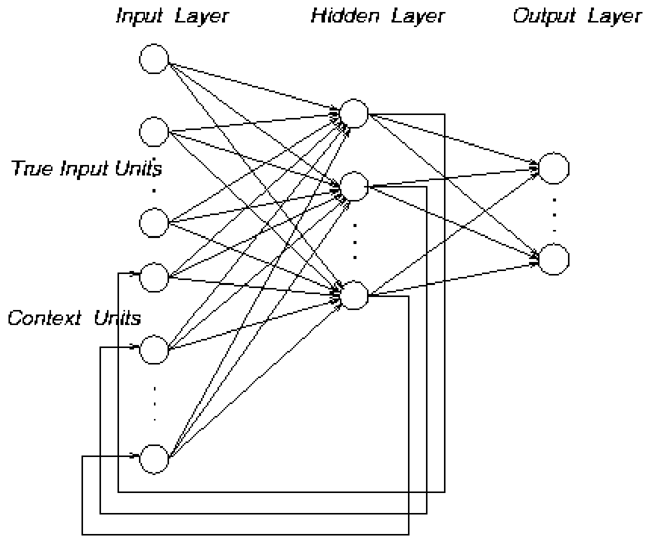
Topologies of Neural Networks



completely connected



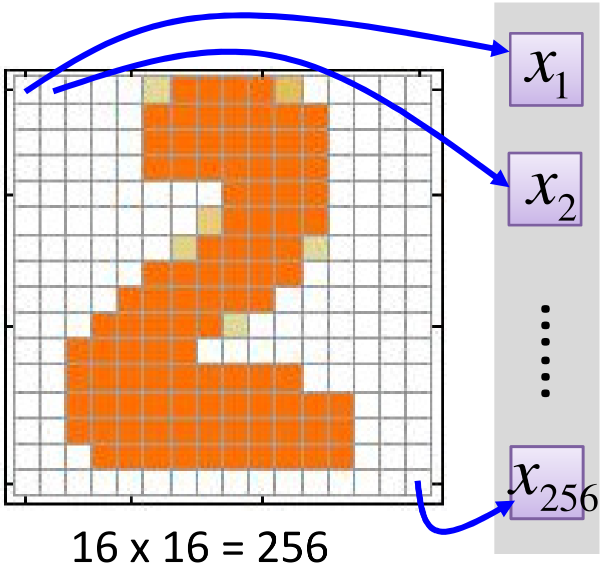
feedforward (directed, a-cyclic)



recurrent (feedback connections)

Handwriting Digit Recognition

Input



x_1

x_2

……

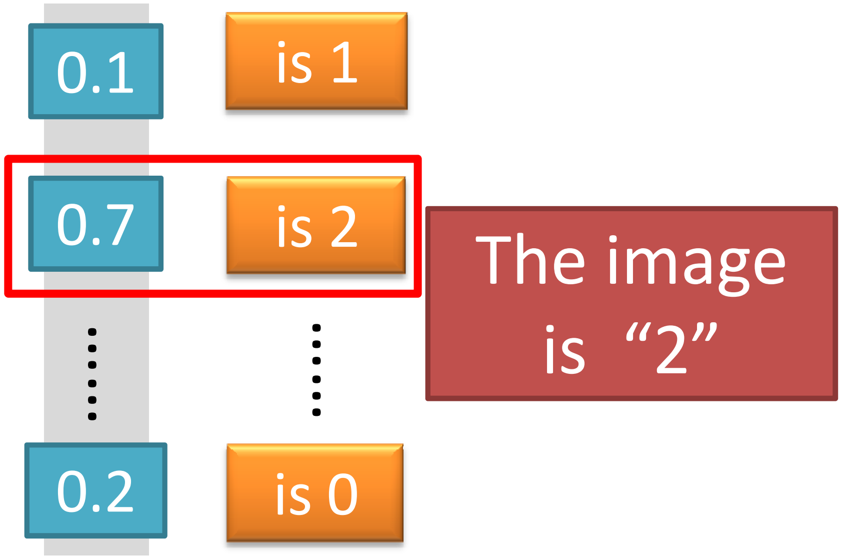
x_{256}

16 x 16 = 256

Ink → 1

No ink → 0

Output



0.1 is 1

0.7 is 2

……

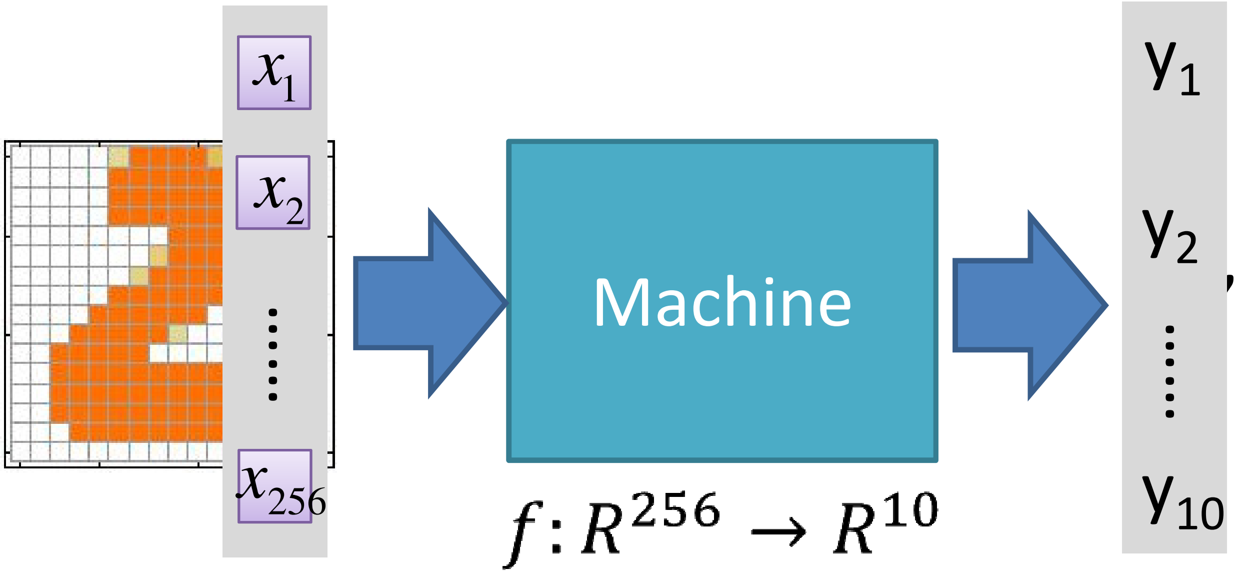
0.2 is 0

The image is “2”

Each dimension represents the confidence of a digit.

Example Application

- Handwriting Digit Recognition



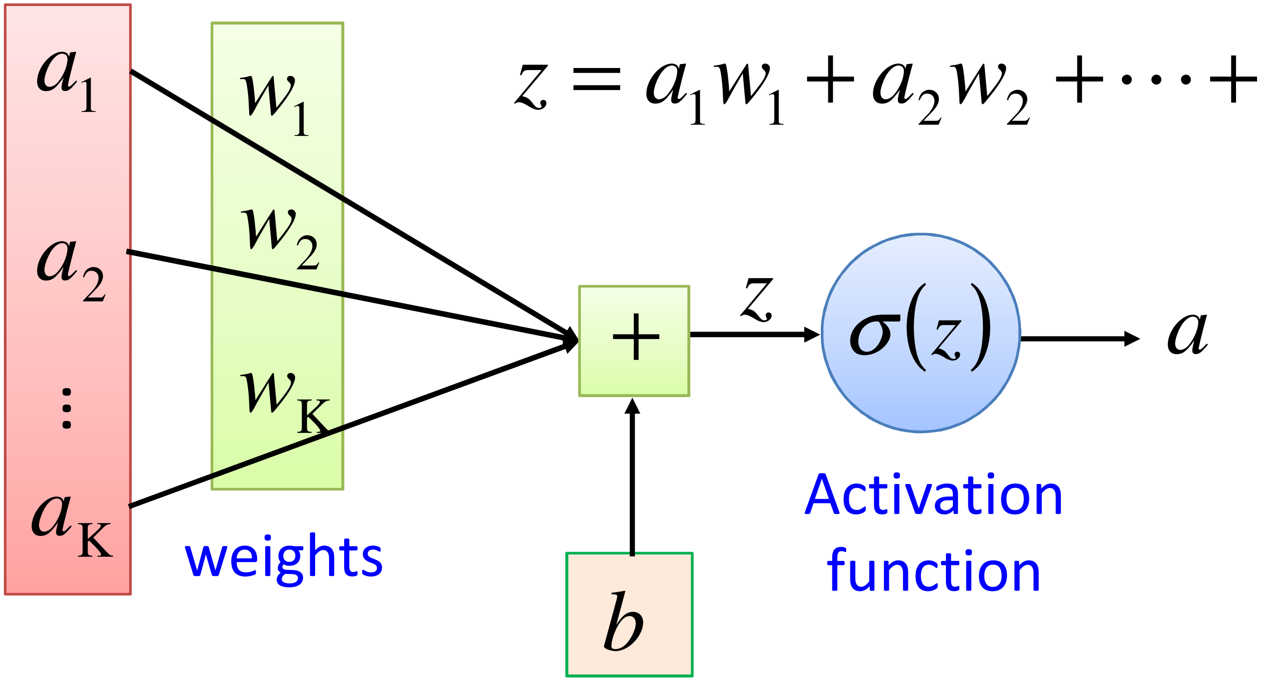
$$f: R^{256} \to R^{10}$$

In deep learning, the function f is represented by neural network

Element of Neural Network

Neuron $f: R^K \to R$

$$z = a_1 w_1 + a_2 w_2 + \cdots + a_K w_K + b$$

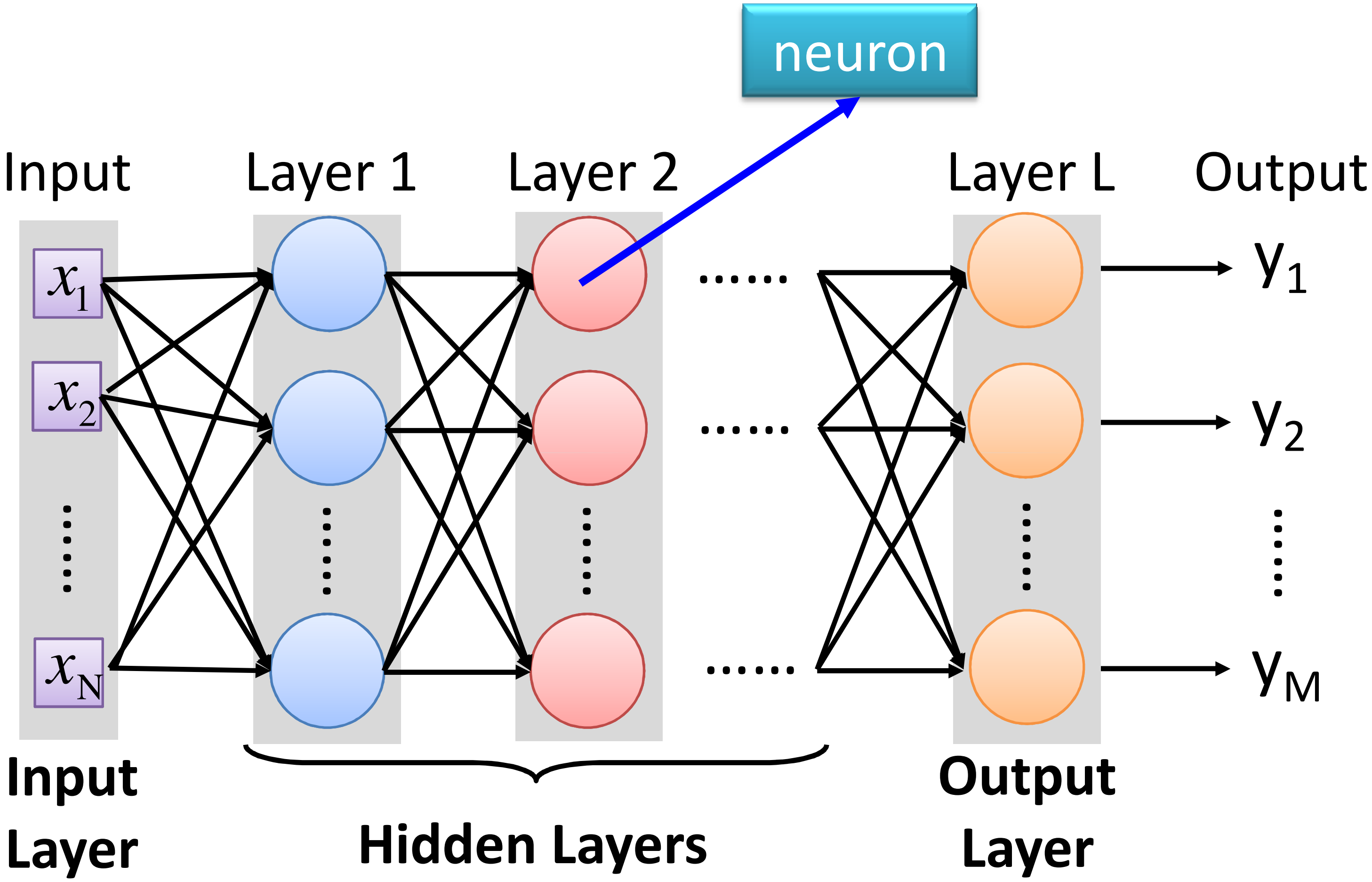


weights

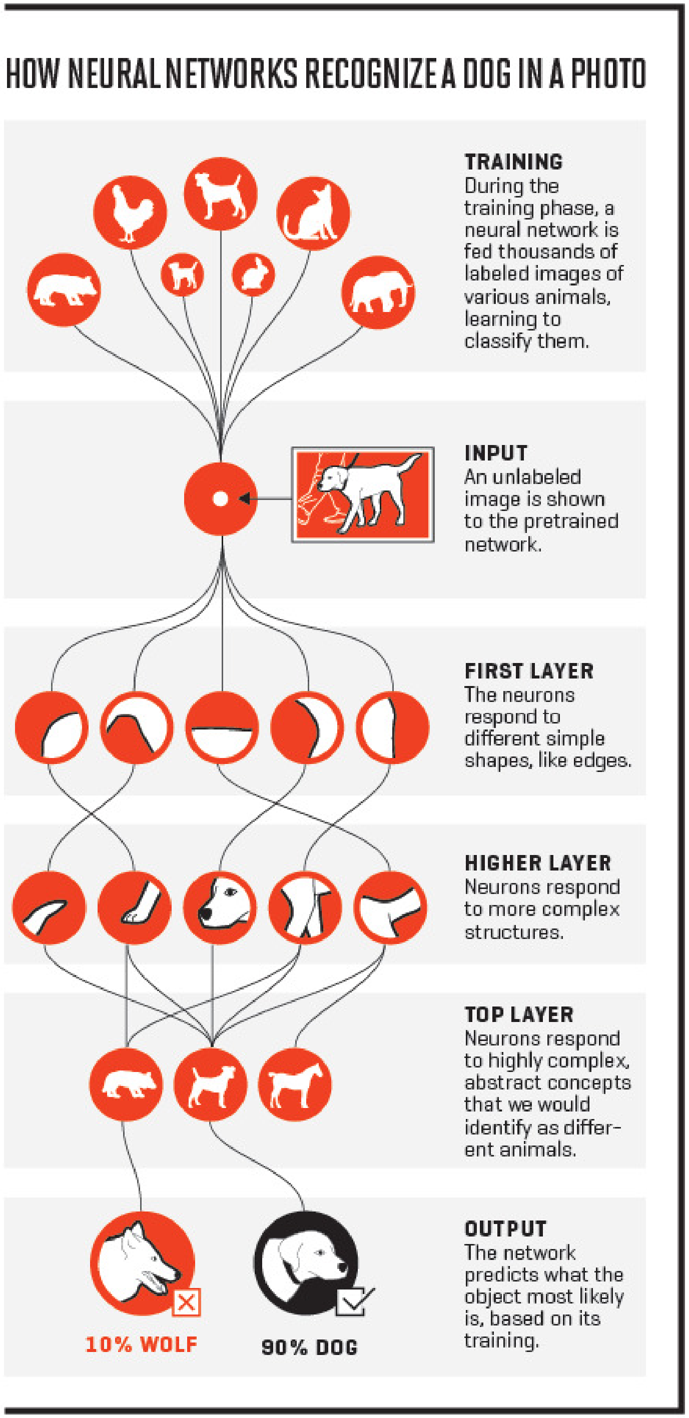
Activation function

bias

Neural Network



Deep means many hidden layers



HOW NEURAL NETWORKS RECOGNIZE A DOG IN A PHOTO

TRAINING
During the training phase, a neural network is fed thousands of labeled images of various animals, learning to classify them.

INPUT
An unlabeled image is shown to the pretrained network.

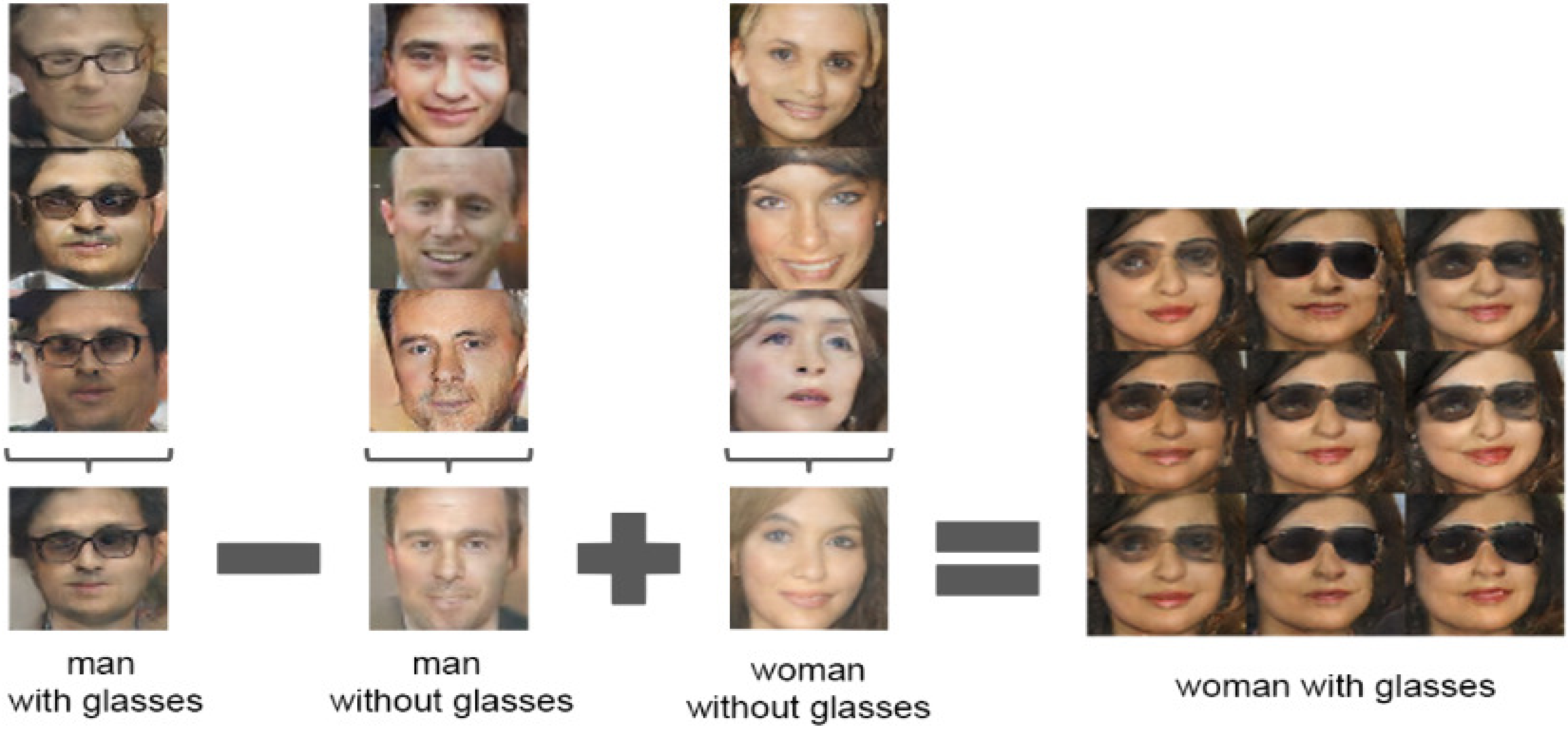
FIRST LAYER
The neurons respond to different simple shapes, like edges.

HIGHER LAYER
Neurons respond to more complex structures.

TOP LAYER
Neurons respond to highly complex, abstract concepts that we would identify as different animals.

OUTPUT
The network predicts what the object most likely is, based on its training.

Latent vectors capture interesting patterns…



Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv:1511.06434 (2015).

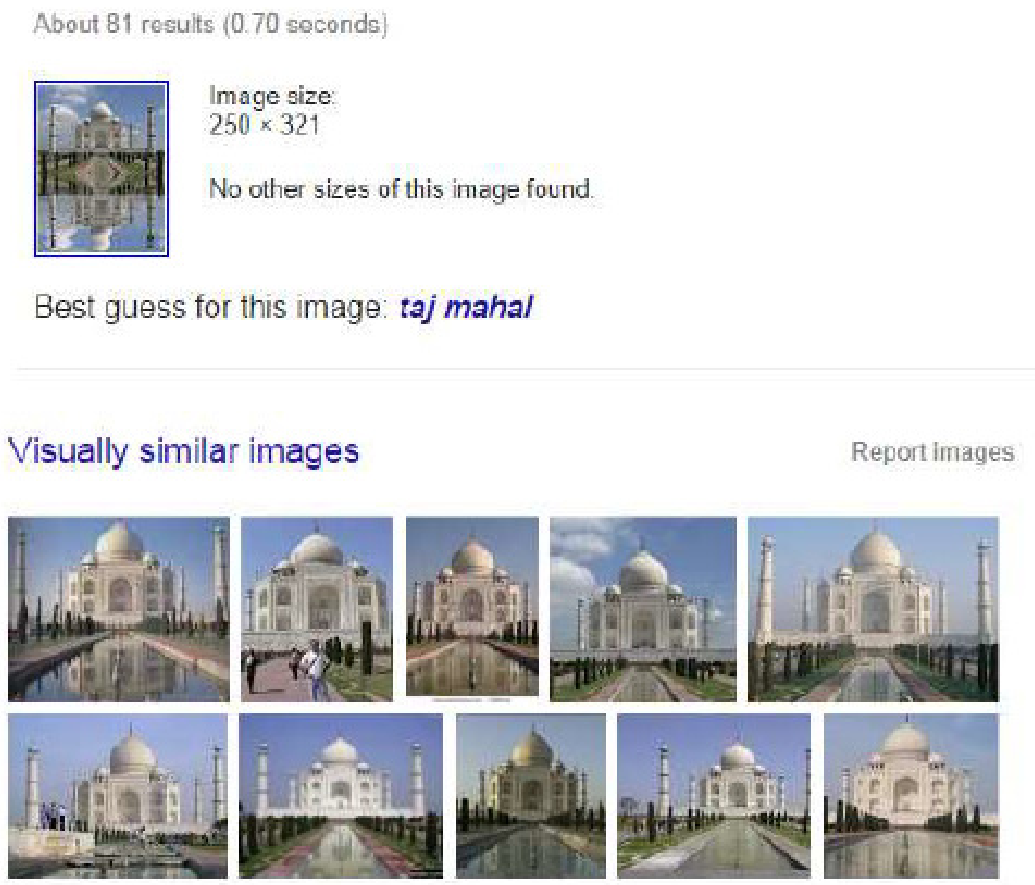
Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging
- [Your favorite area]

Applications (conti..)

- Spam Email Detection
- Machine Translation (Language Translation)
- Image Search (Similarity)
- Clustering (KMeans) : Amazon
- Recommendations
- Classification : Google News
- Text Summarization - Google News
- Rating a Review/Comment: Yelp
- Fraud detection : Credit card Providers
- Decision Making : e.g. Bank/Insurance sector
- Sentiment Analysis
- Speech Understanding – iPhone with Siri
- Face Detection – Facebook’s Photo tagging

Similar/Duplicate Images



About 81 results (0.70 seconds)

Image size:
250 × 321

No other sizes of this image found.

Best guess for this image: ***taj mahal***

Visually similar images

Report images

Credit: https://www.google.co.in/

Remember

Features ?

(Feature Extraction)

Can be :

- Width
- Height
- Contrast
- Brightness
- Position
- Hue
- Colors

Check this :

LIRE (Lucene Image REtrieval) library - https://code.google.com/p/lire/

Popular Frameworks/Tools

- Weka
- Carrot2
- Gate
- OpenNLP
- LingPipe
- Stanford NLP
- Mallet – Topic Modelling
- Gensim – Topic Modelling (Python)
- Apache Mahout
- MLib – Apache Spark
- scikit-learn - Python
- LIBSVM : Support Vector Machines
- and many more…

AI APPLICATIONS

Image Classification

Object Detection

COMPUTER VISION

Voice Recognition

Language Translation

SPEECH & AUDIO

Recommendation Engines

Sentiment Analysis

NATURAL LANGUAGE PROCESSING

greatlearning
Learning for Life

20 DEEP LEARNING Applications

1 Self Driving Cars

2 Entertainment

3 Visual Recognition

4 Virtual Assistants

5 Fraud Detection

6 Natural Language Processing

7 News Aggregation and Fraud News Detection

8 Detecting Developmental Delay in Children

9 Colourisation of Black and White images

10 Adding sounds to silent movies

Healthcare 11

Personalisations 12

Automatic Machine Translation 13

Automatic Handwriting Generation 14

Demographic & Election Predictions 15

16 Automatic Game Playing

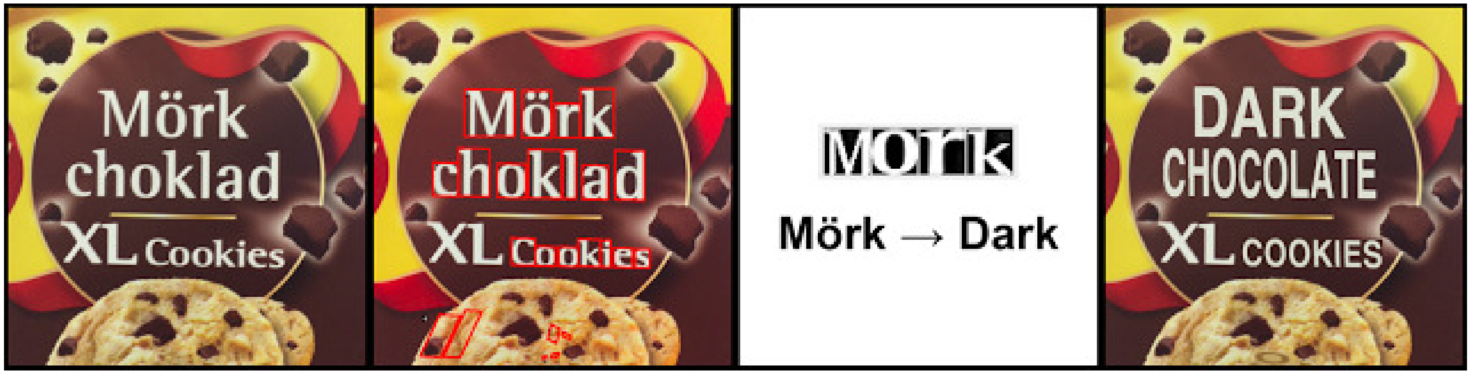
17 Language Translations

18 Pixel Restoration

19 Photo Descriptions

20 Deep Dreaming

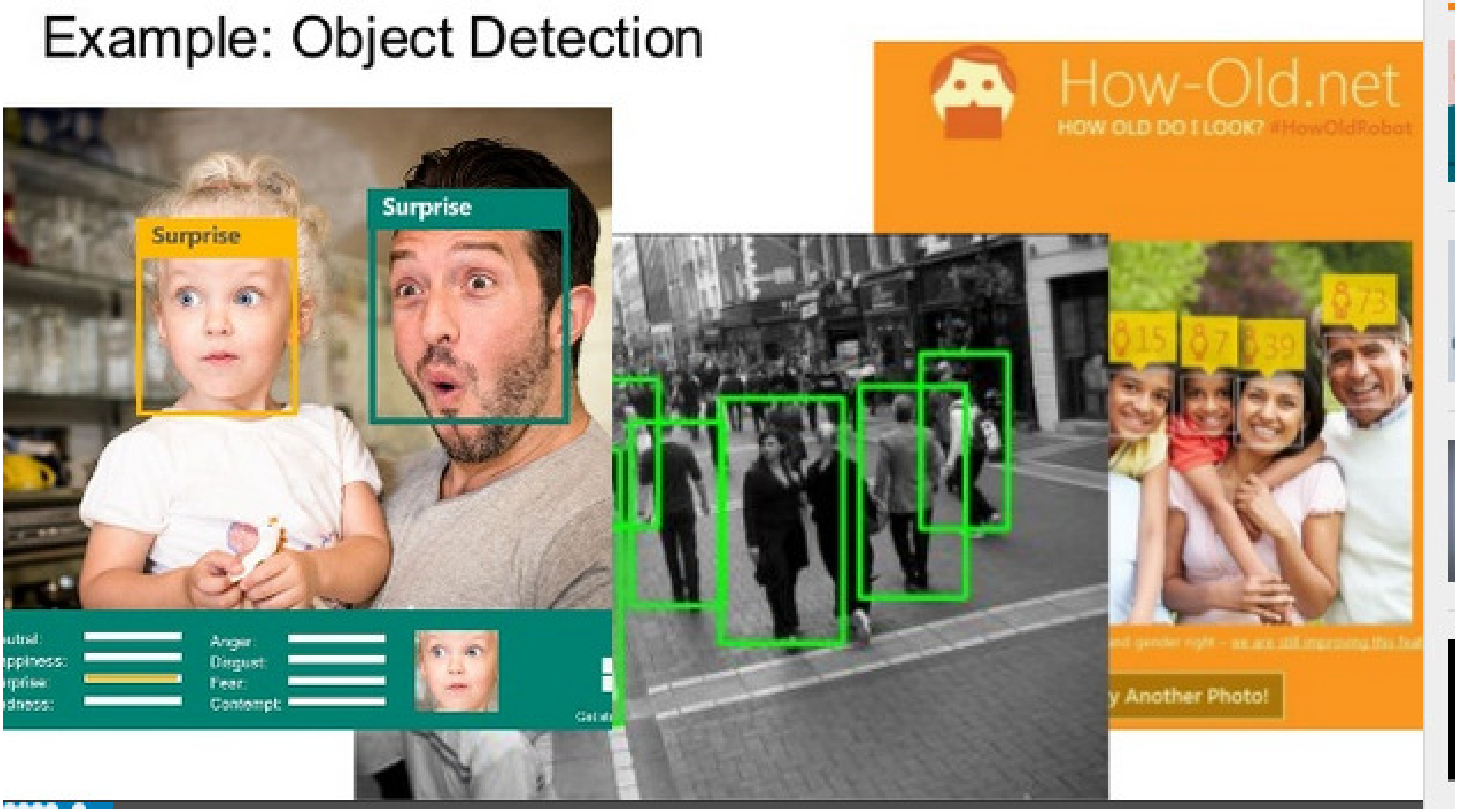
Image Translation



Applications

- Deep Learning AI is revolutionizing the filmmaking process as cameras learn to study human body language to imbibe in virtual characters.
- A deep learning model tends to associate the video frames with a database of pre-recorded sounds to select appropriate sounds for the scene
- https://youtu.be/0FW99AQmMc8
- http://news.mit.edu/2016/artificial-intelligence-produces-realistic-sounds-0613

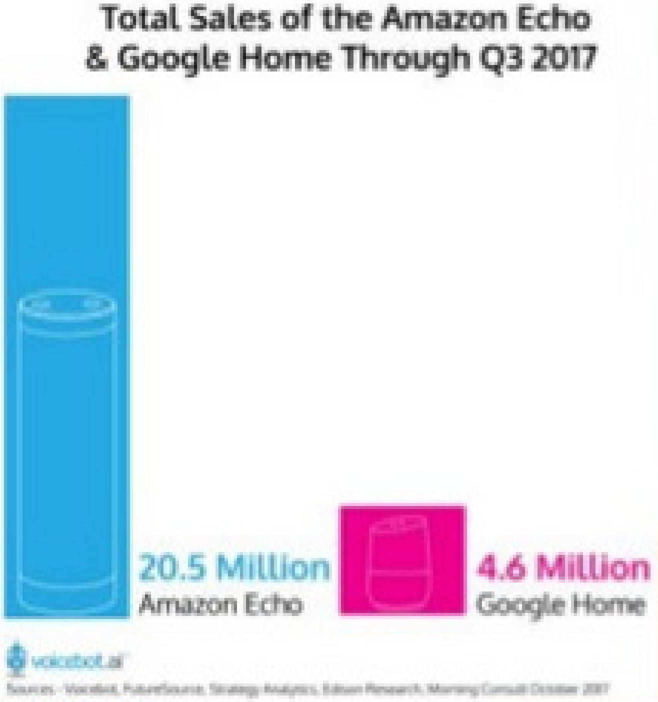
Example: Object Detection



Case: Amazon Echo

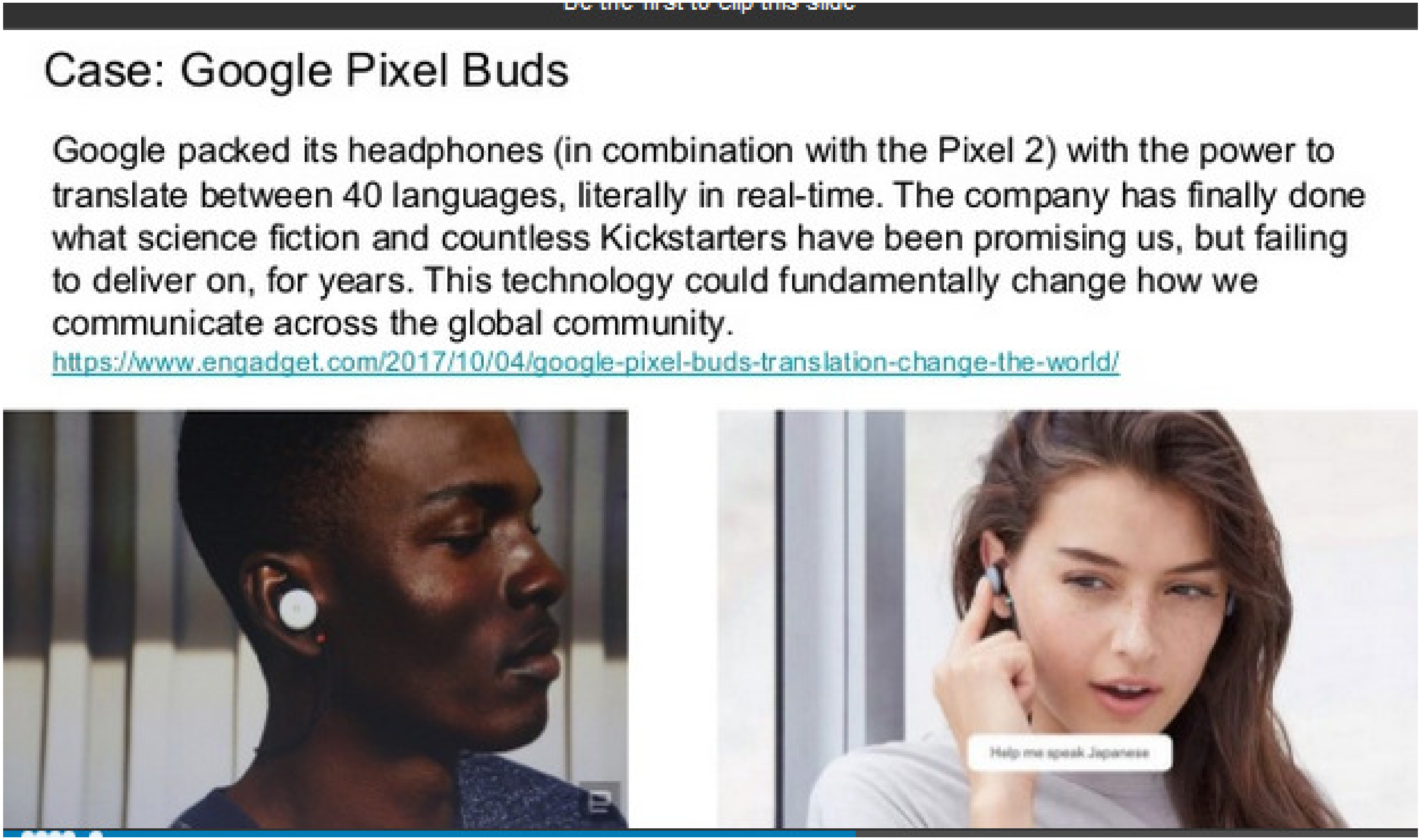
Amazon Alexa is in more than 20 million devices. The vast majority of these are in the Amazon Echo portfolio.

https://www.voicebot.ai/2017/10/27/bezos-says-20-million-amazon-alexa-devices-sold/



Total Sales of the Amazon Echo & Google Home Through Q3 2017

Sources - Voicebot, FutureSource, Strategy Analytics, Edison Research, Morning Consult October 2017



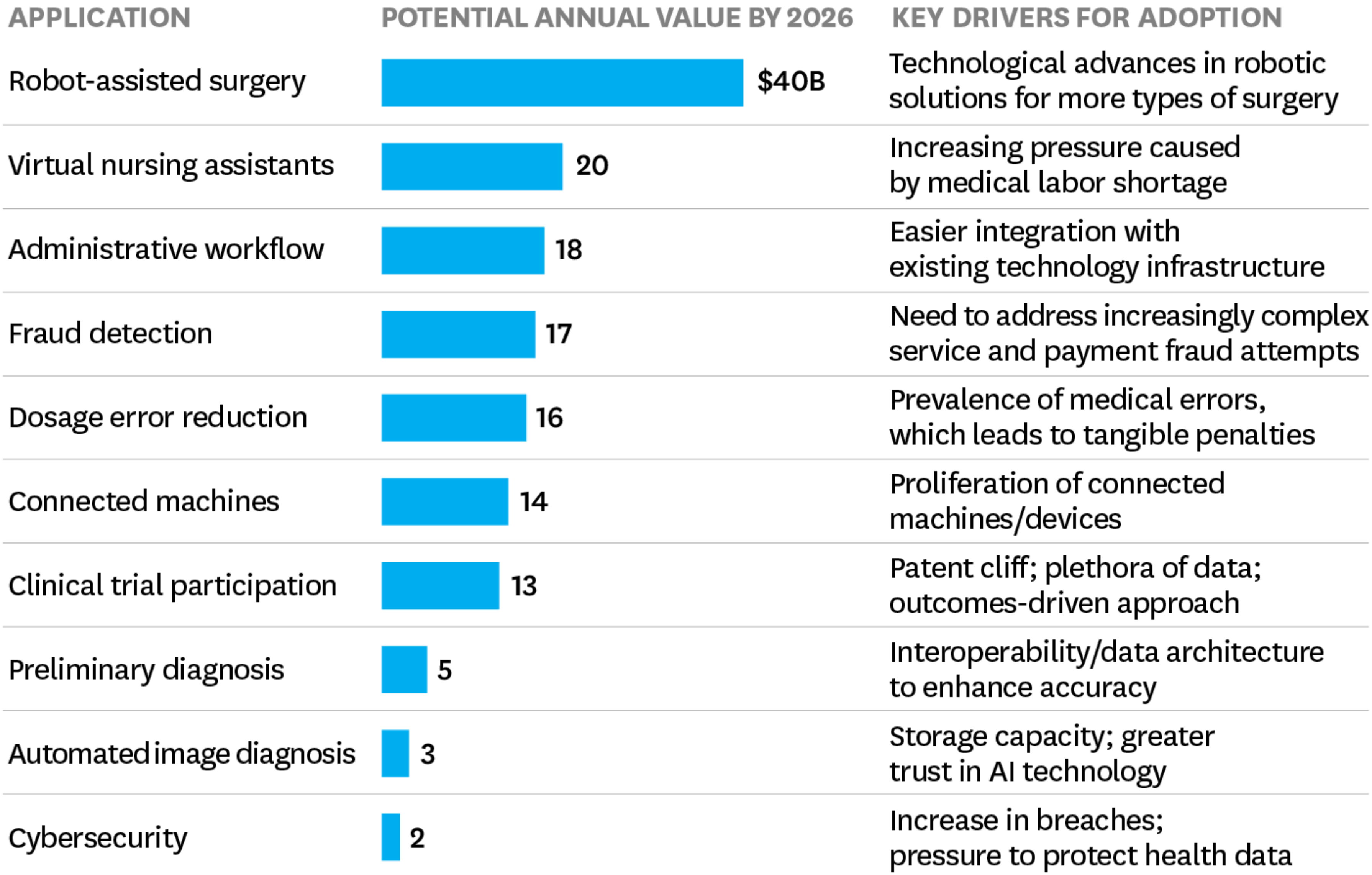
Be the first to clip this slide

Case: Google Pixel Buds

Google packed its headphones (in combination with the Pixel 2) with the power to translate between 40 languages, literally in real-time. The company has finally done what science fiction and countless Kickstarters have been promising us, but failing to deliver on, for years. This technology could fundamentally change how we communicate across the global community.

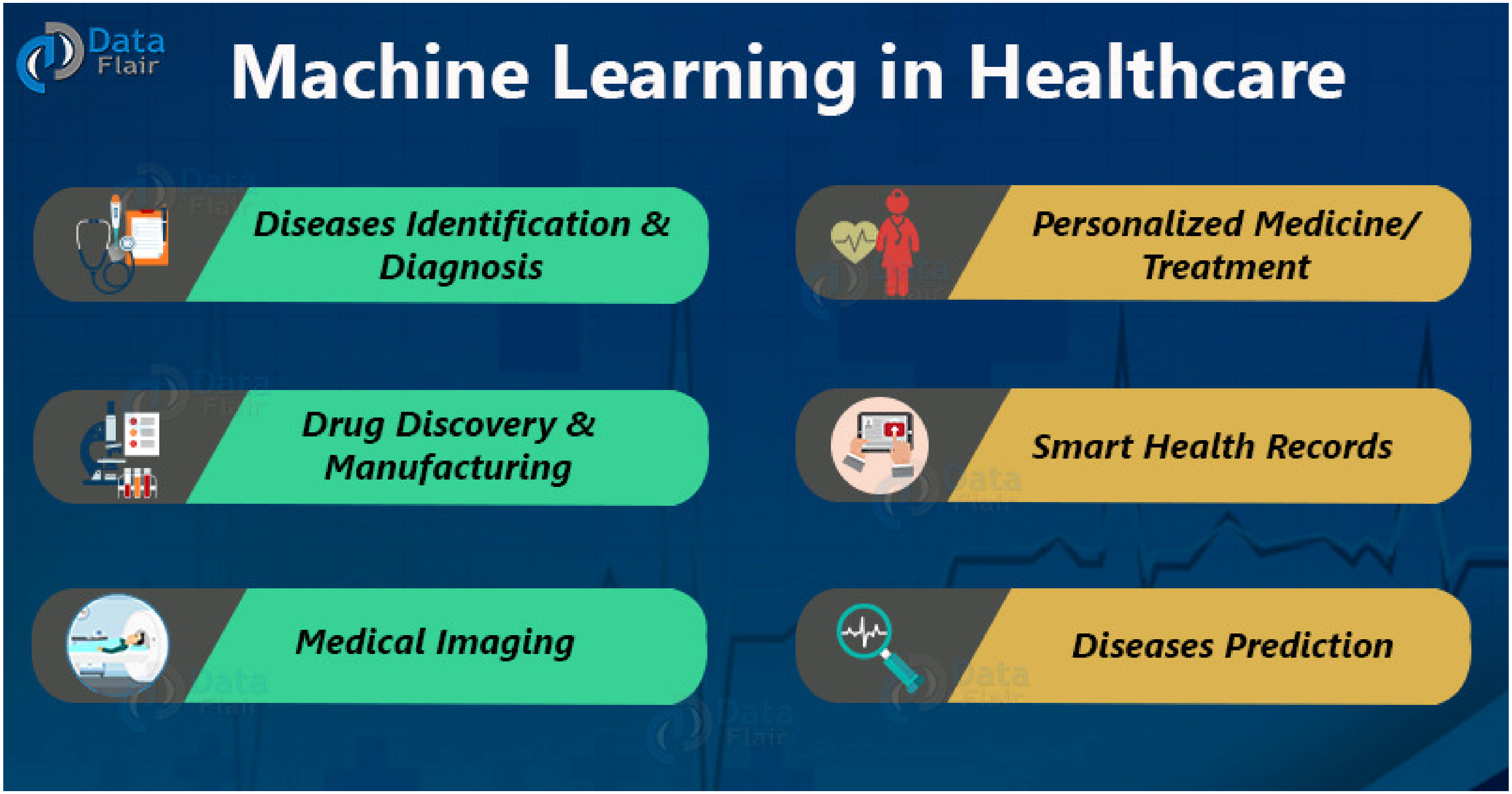
https://www.engadget.com/2017/10/04/google-pixel-buds-translation-change-the-world/

10 AI Applications That Could Change Health Care



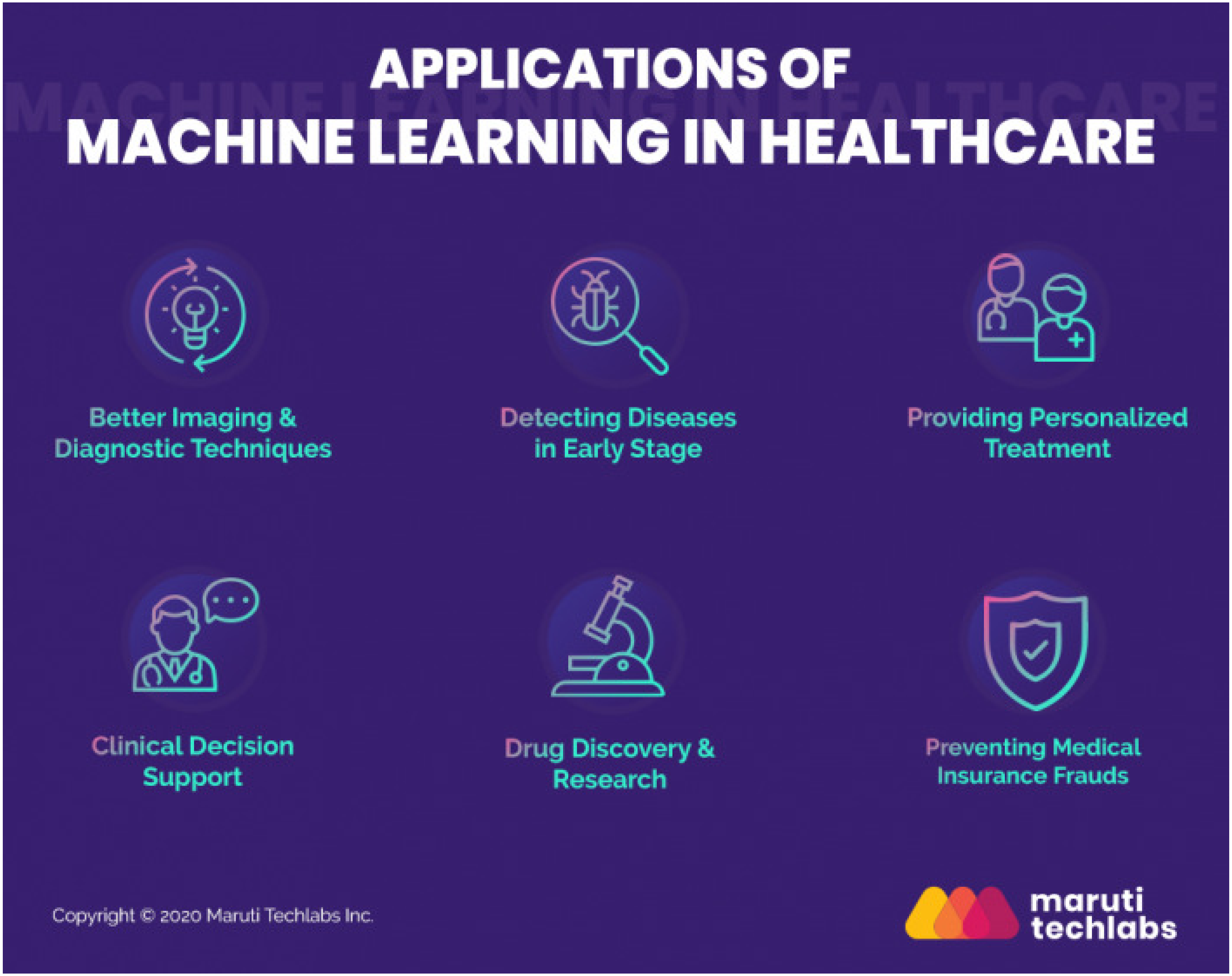
APPLICATION	POTENTIAL ANNUAL VALUE BY 2026	KEY DRIVERS FOR ADOPTION
Robot-assisted surgery	$40B	Technological advances in robotic solutions for more types of surgery
Virtual nursing assistants	20	Increasing pressure caused by medical labor shortage
Administrative workflow	18	Easier integration with existing technology infrastructure
Fraud detection	17	Need to address increasingly complex service and payment fraud attempts
Dosage error reduction	16	Prevalence of medical errors, which leads to tangible penalties
Connected machines	14	Proliferation of connected machines/devices
Clinical trial participation	13	Patent cliff; plethora of data; outcomes-driven approach
Preliminary diagnosis	5	Interoperability/data architecture to enhance accuracy
Automated image diagnosis	3	Storage capacity; greater trust in AI technology
Cybersecurity	2	Increase in breaches; pressure to protect health data

SOURCE ACCENTURE



Machine Learning in Healthcare

- Diseases Identification & Diagnosis
- Personalized Medicine/ Treatment
- Drug Discovery & Manufacturing
- Smart Health Records
- Medical Imaging
- Diseases Prediction

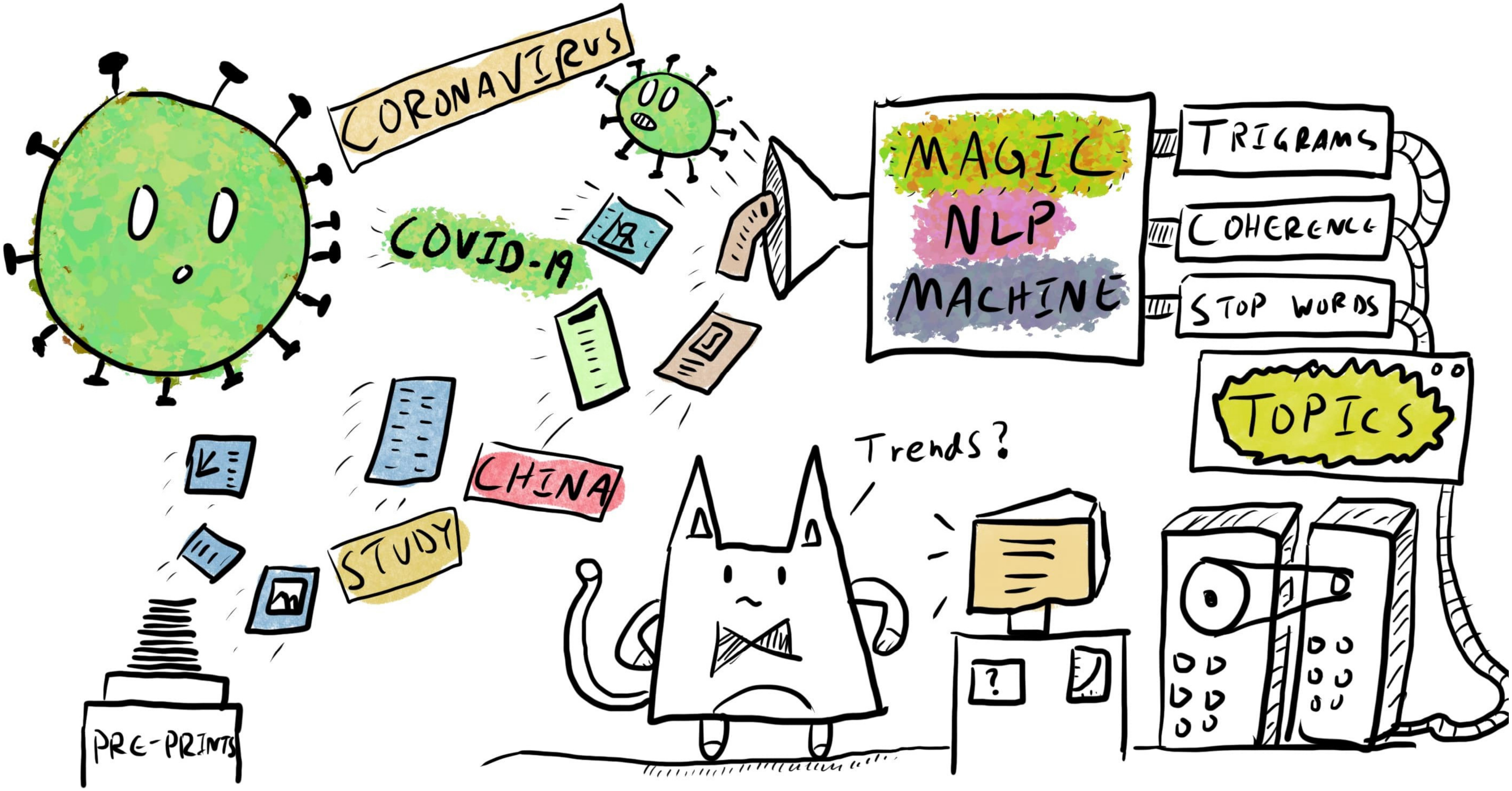


APPLICATIONS OF MACHINE LEARNING IN HEALTHCARE

- Better Imaging & Diagnostic Techniques
- Detecting Diseases in Early Stage
- Providing Personalized Treatment
- Clinical Decision Support
- Drug Discovery & Research
- Preventing Medical Insurance Frauds

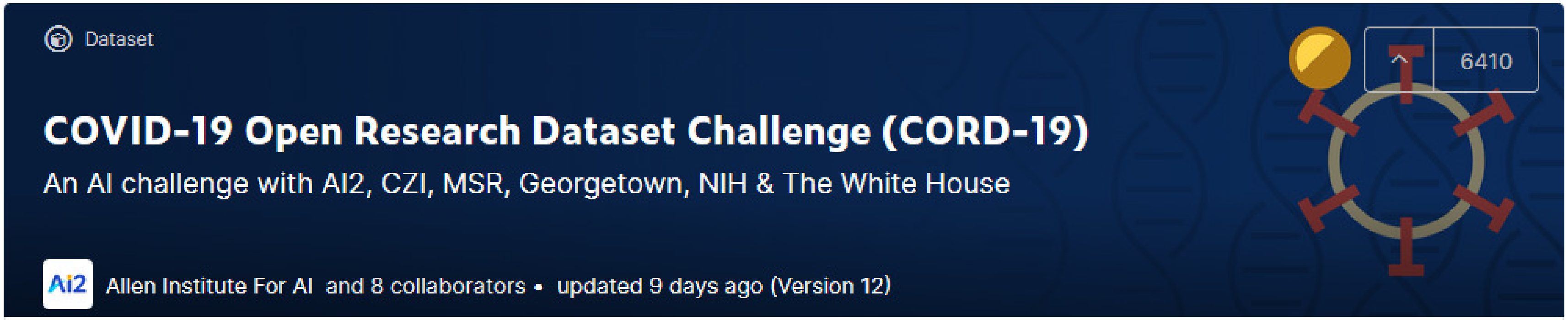
Predict A Doctor's Consultation Fee Hackathon

Bio-NLP



https://towardsdatascience.com/summarising-the-latest-research-on-coronavirus-with-nlp-and-topic-modelling-28b867ad9860

Covid NLP datasets

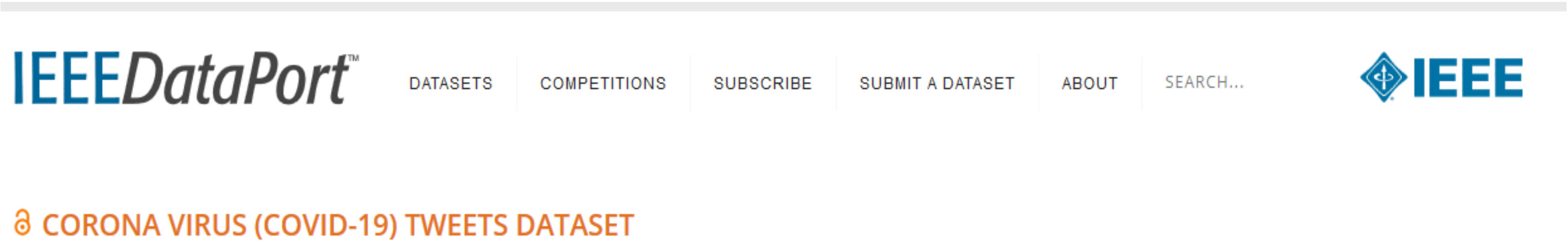


COVID-19 Open Research Dataset Challenge (CORD-19)

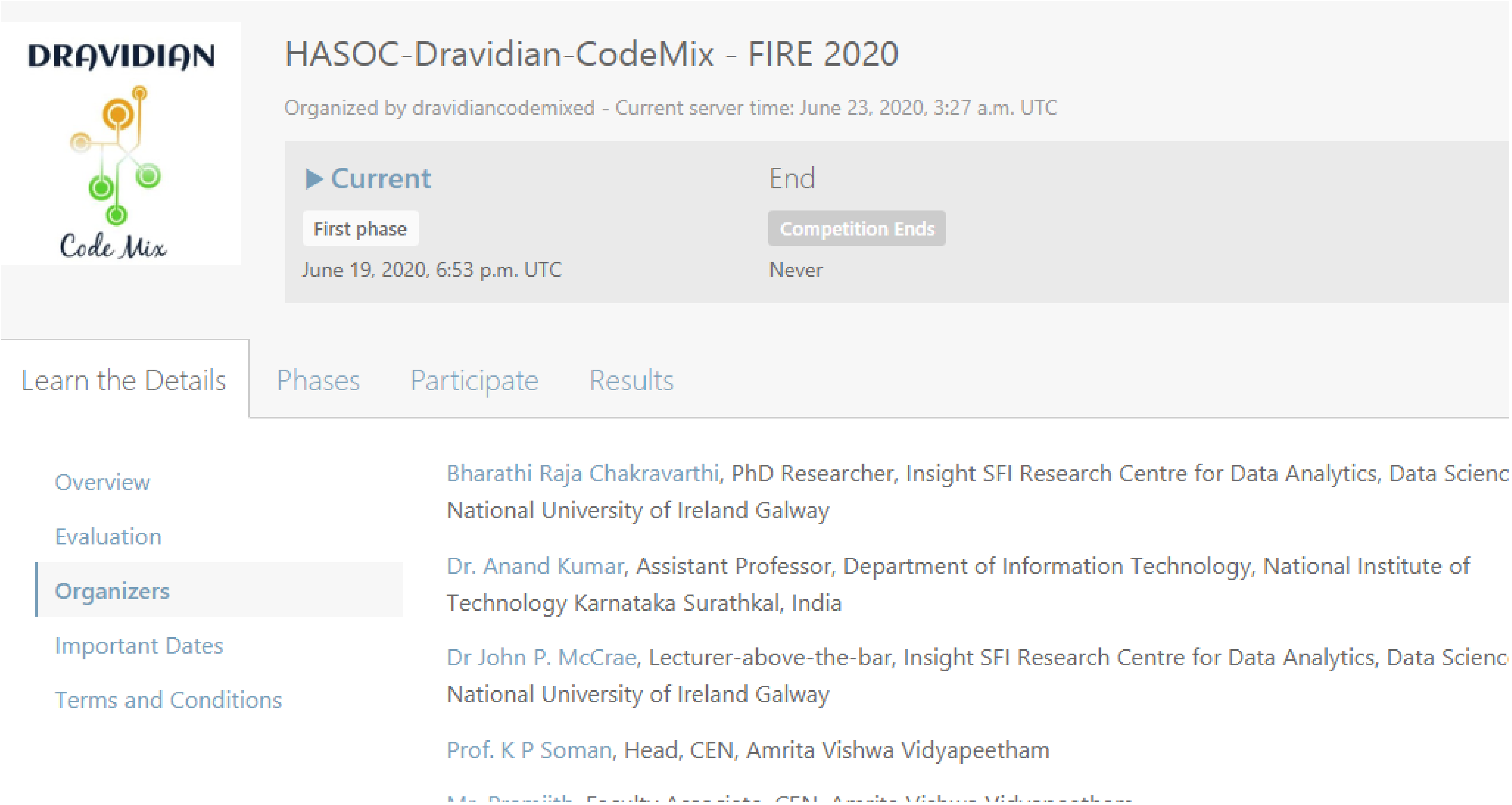
An AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House

Allen Institute For AI and 8 collaborators • updated 9 days ago (Version 12)

Allen Institute for AI Open Research Dataset (CORD-19), over 47,000 scholarly articles, including over 36,000 with full text, about COVID-19 and the coronavirus family of viruses for use by the global research community.



CORONA VIRUS (COVID-19) TWEETS DATASET



HASOC-Dravidian-CodeMix - FIRE 2020

Organized by dravidiancodemixed - Current server time: June 23, 2020, 3:27 a.m. UTC

June 19, 2020, 6:53 p.m. UTC

Never

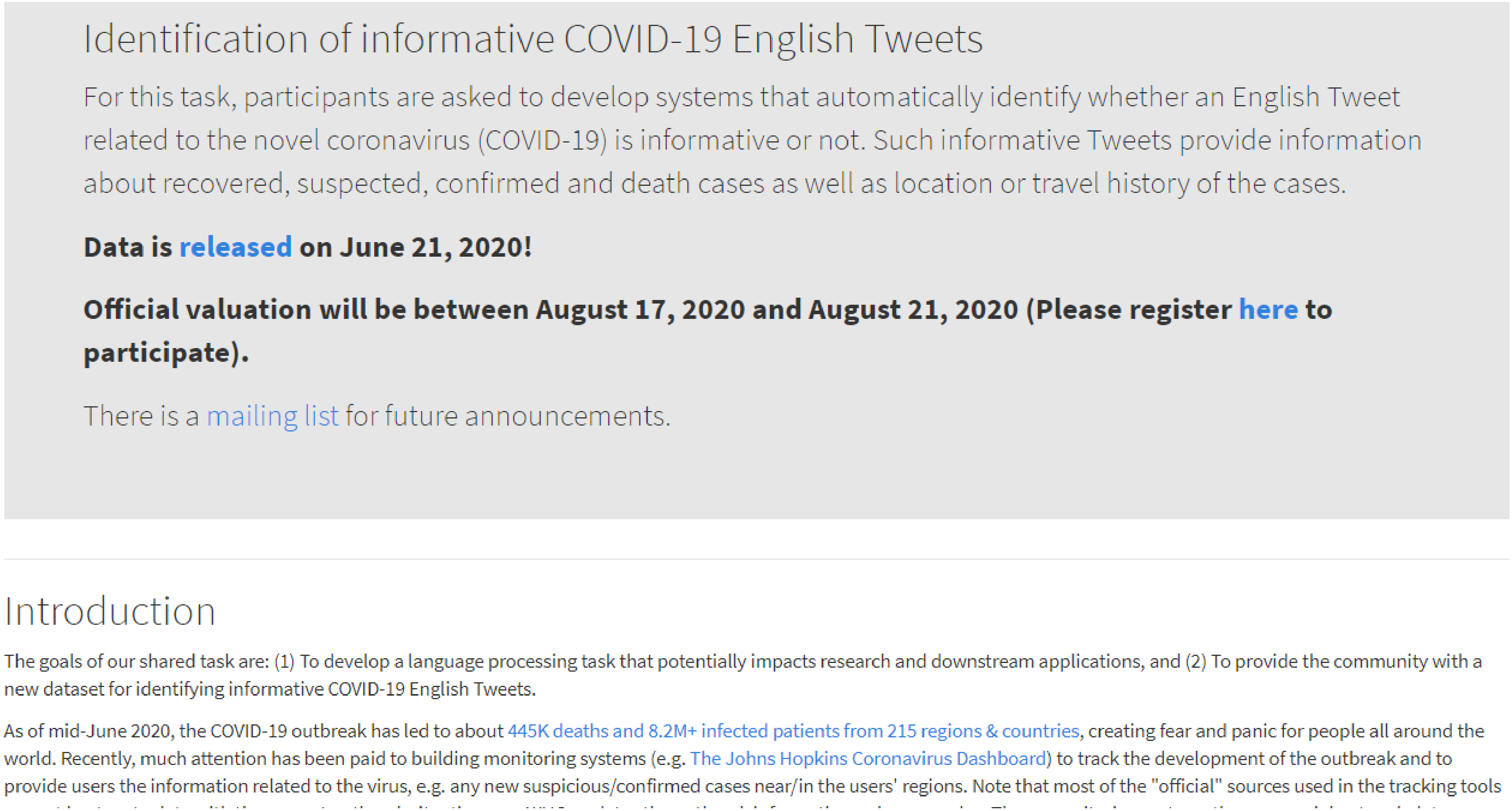
Bharathi Raja Chakravarthi, PhD Researcher, Insight SFI Research Centre for Data Analytics, Data Scienc
National University of Ireland Galway

Dr. Anand Kumar, Assistant Professor, Department of Information Technology, National Institute of Technology Karnataka Surathkal, India

Dr John P. McCrae, Lecturer-above-the-bar, Insight SFI Research Centre for Data Analytics, Data Scienc
National University of Ireland Galway

Prof. K P Soman, Head, CEN, Amrita Vishwa Vidyapeetham

Mr. Premjith, Faculty Associate, CEN, Amrita Vishwa Vidyapeetham



Identification of informative COVID-19 English Tweets

For this task, participants are asked to develop systems that automatically identify whether an English Tweet related to the novel coronavirus (COVID-19) is informative or not. Such informative Tweets provide information about recovered, suspected, confirmed and death cases as well as location or travel history of the cases.

Data is released on June 21, 2020!

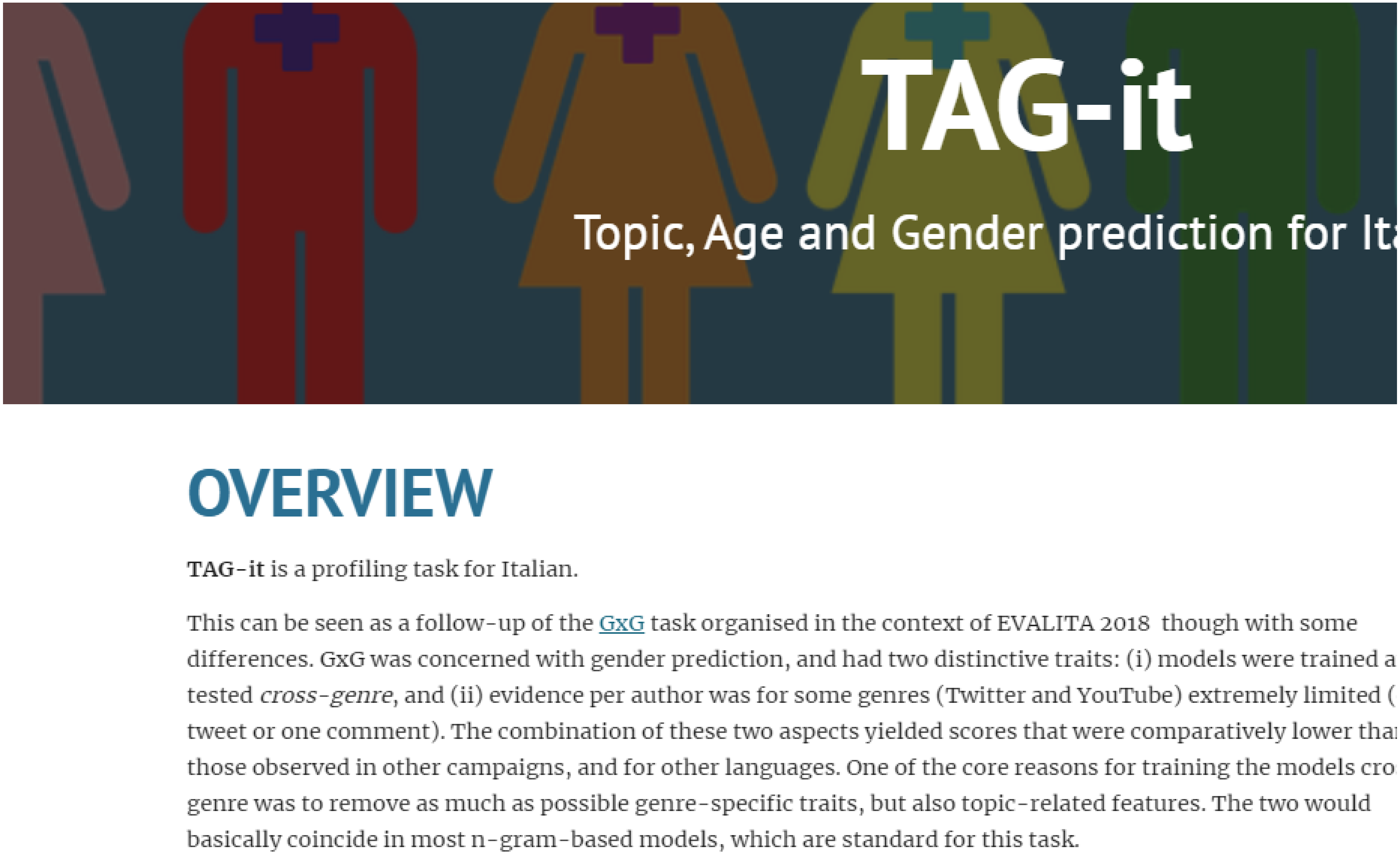
Official valuation will be between August 17, 2020 and August 21, 2020 (Please register here to participate).

There is a mailing list for future announcements.

Introduction

The goals of our shared task are: (1) To develop a language processing task that potentially impacts research and downstream applications, and (2) To provide the community with a new dataset for identifying informative COVID-19 English Tweets.

As of mid-June 2020, the COVID-19 outbreak has led to about 445K deaths and 8.2M+ infected patients from 215 regions & countries, creating fear and panic for people all around the world. Recently, much attention has been paid to building monitoring systems (e.g. The Johns Hopkins Coronavirus Dashboard) to track the development of the outbreak and to provide users the information related to the virus, e.g. any new suspicious/confirmed cases near/in the users' regions. Note that most of the "official" sources used in the tracking tools



TAG-it

Topic, Age and Gender prediction for It

OVERVIEW

TAG-it is a profiling task for Italian.

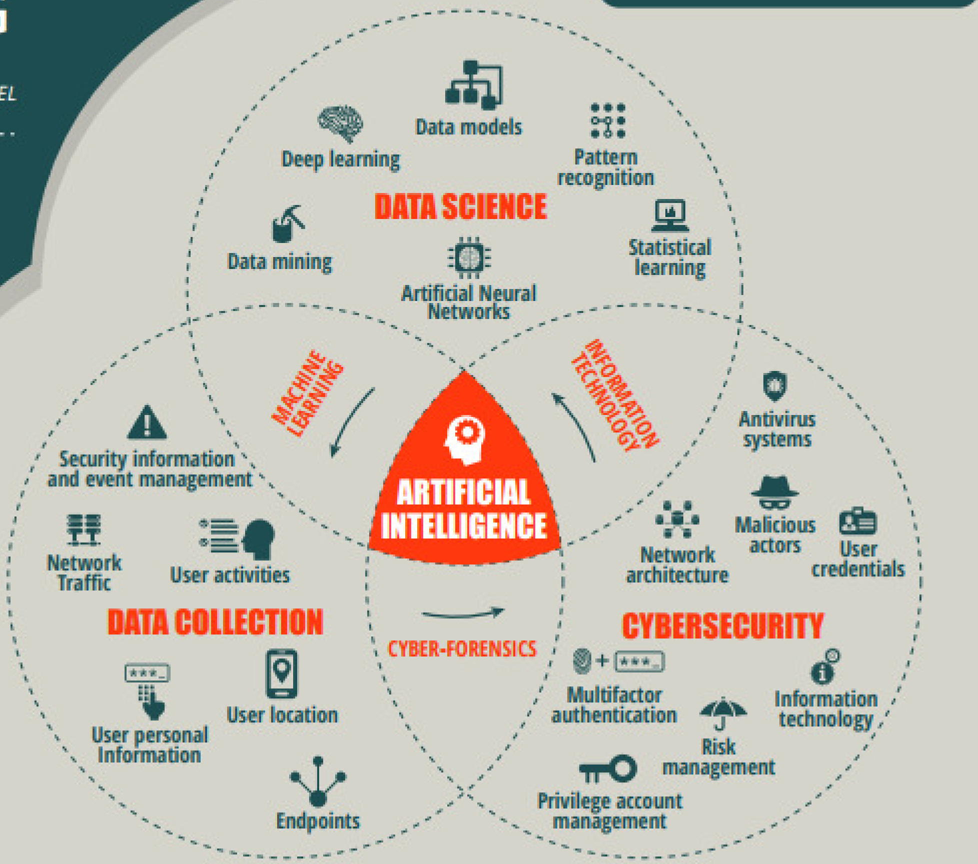
This can be seen as a follow-up of the GxG task organised in the context of EVALITA 2018 though with some differences. GxG was concerned with gender prediction, and had two distinctive traits: (i) models were trained a tested *cross-genre*, and (ii) evidence per author was for some genres (Twitter and YouTube) extremely limited (tweet or one comment). The combination of these two aspects yielded scores that were comparatively lower tha those observed in other campaigns, and for other languages. One of the core reasons for training the models cro genre was to remove as much as possible genre-specific traits, but also topic-related features. The two would basically coincide in most n-gram-based models, which are standard for this task.

WHAT CAN MACHINE LEARNING DO FOR CYBERSECURITY?

A POTENT NEW ARSENAL FOR IT AND CYBERSECURITY PERSONNEL

- User entity behavioral analytics, deep learning, automation
- Assist IT professionals and defend against new cyberthreats
- Better predictive models, lower FPR, distill new metrics
- Fraud and anomaly detection
- Defend against new cyberthreats
- Better use of internal data and global repositories
- Tackle device influx and enhanced data loss prevention (DLP) solutions

Analytics and Forensics



DATA SCIENCE

Deep learning

Data models

Pattern recognition

Data mining

Statistical learning

Artificial Neural Networks

MACHINE LEARNING

INFORMATION TECHNOLOGY

ARTIFICIAL INTELLIGENCE

DATA COLLECTION

Security information and event management

Network Traffic

User activities

User personal Information

User location

Endpoints

CYBER-FORENSICS

CYBERSECURITY

Antivirus systems

Malicious actors

Network architecture

User credentials

Multifactor authentication

Information technology

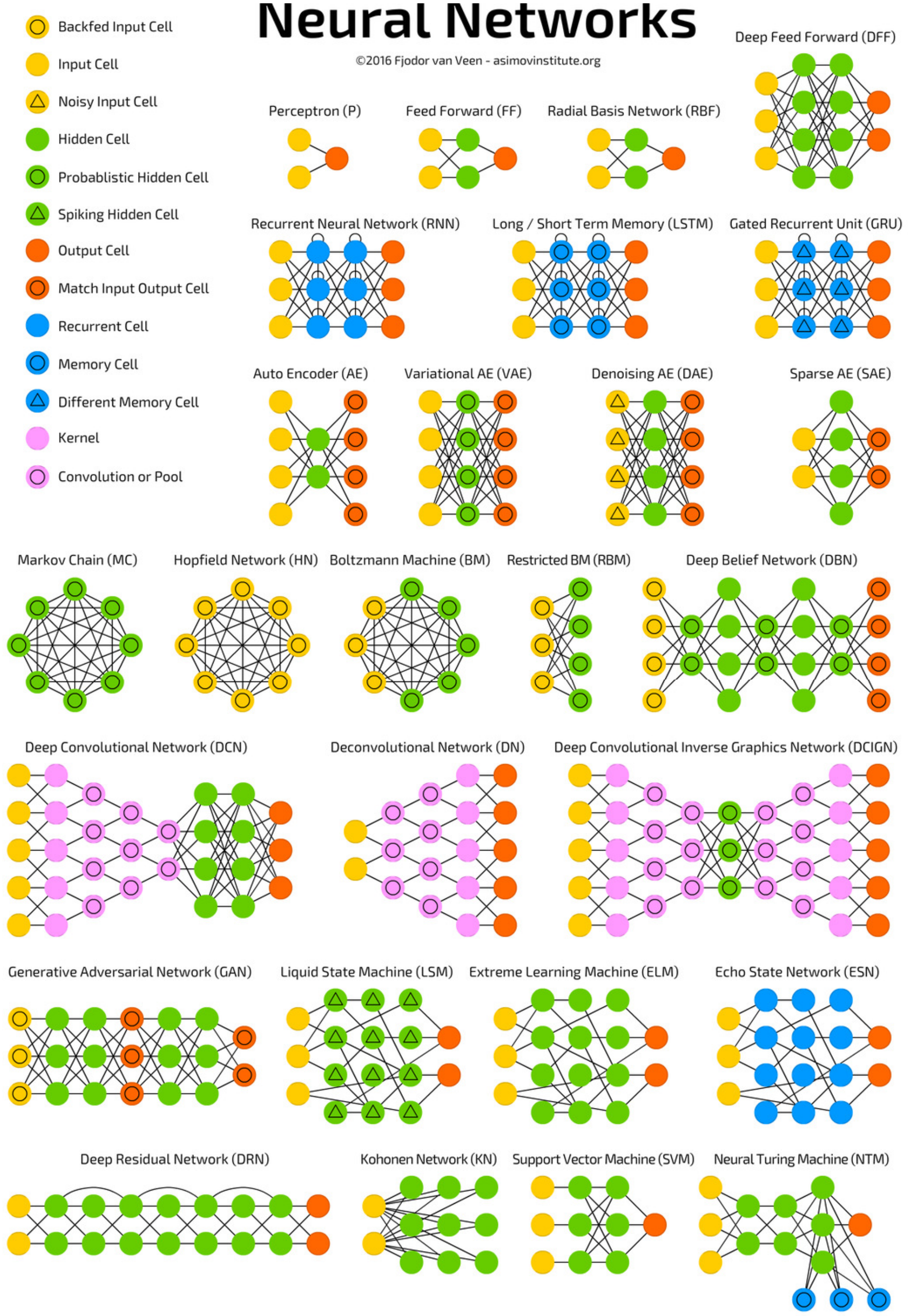
Risk management

Privilege account management

- **Data Science**: Applying machine learning and creating new data models to combat new threats
- **Data Collection**: Harnessing the power of data from a wide spectrum of sources
- **Cybersecurity**: Domain-specific knowledge and versatility in an ever-changing environment

Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org



- http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
- https://www.mygreatlearning.com/blog/deep-learning-applications/
- http://www.r2d3.us/visual-intro-to-machine-learning-part-2/
- http://www.hpc.lsu.edu/training/weekly-materials/2016-Fall/machine_learning_qb2_fall_2016.pdf
- https://vas3k.com/blog/machine_learning/