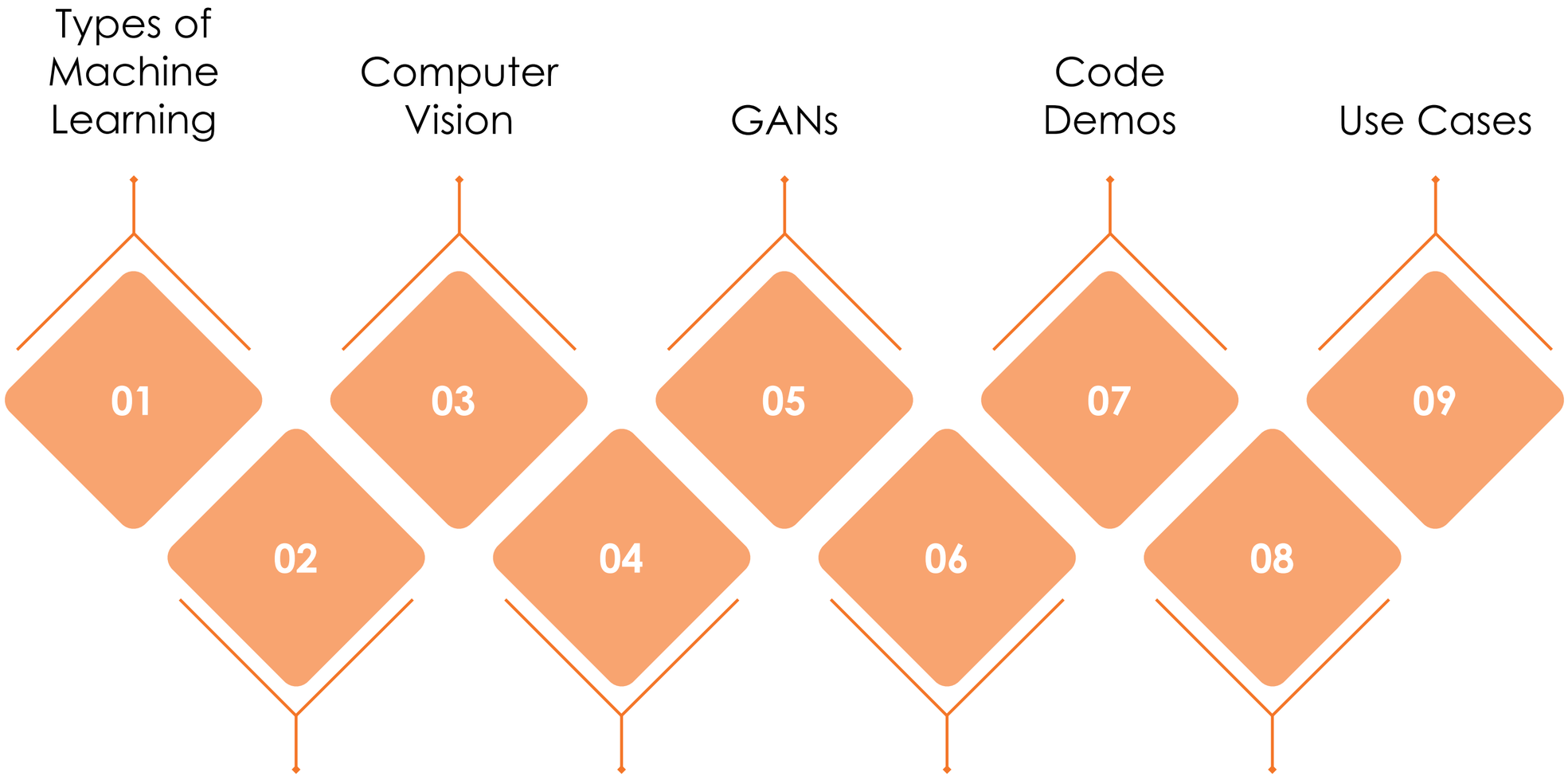
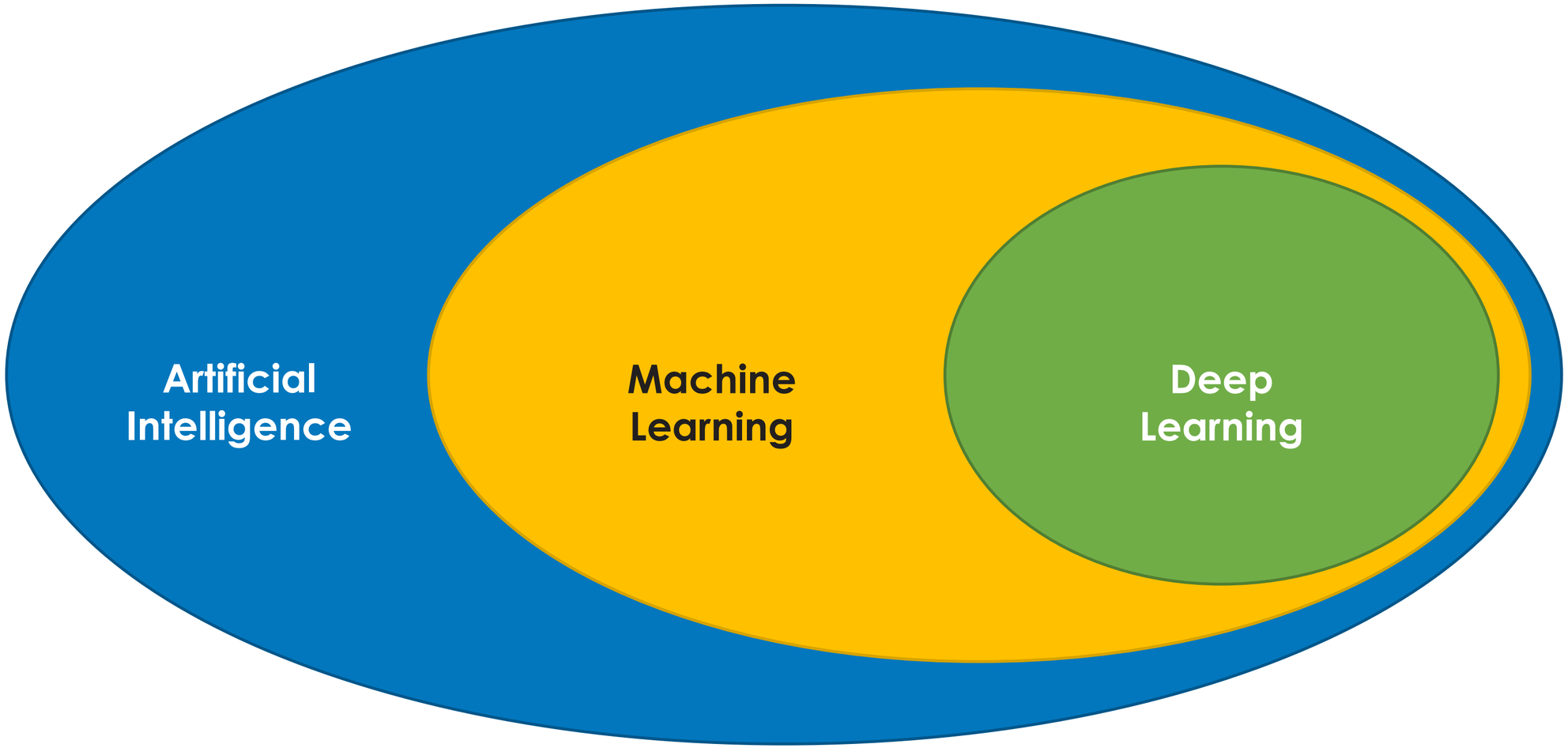
BUILDING INTELLIGENT SYSTEMS

PETER MORGAN

Table of contents



Machine Learning - The Big Picture



Types of Machine Learning

- Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed

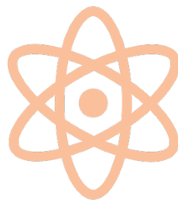
- Machine learning algorithms can be broadly classified into three categories – supervised, unsupervised and reinforcement learning

- Supervised learning
    - Labelled data

- Unsupervised learning
    - Unlabelled data

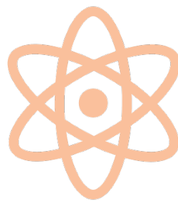
- Reinforcement learning
    - Reward/penalty system

- Also, AGI
    - Artificial General Intelligence (holy grail)



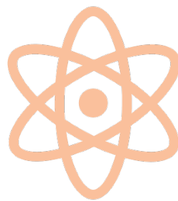
Supervised Learning Algorithms

- Regression
- Classification
- K nearest neighbors (kNN)
- Support Vector Machines (SVM)
- Decision Trees
- Random Forests
- Boosting
- Hidden Markov Models



Supervised Learning Algorithms

- Generally, supervised learning algorithms can be divided into two categories
- Regression
  - Fitting a line, curve (polynomial) or plane (hyperplane) through the data
- Classification
  - Classifying n-dimensional data points into different categories
- Some address both regression and classification



Unsupervised Learning Algorithms

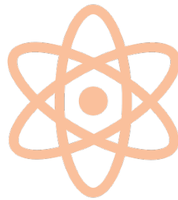
- Clustering - cluster analysis is the assignment of a set of observations into subsets called clusters, so that observations within the same cluster are similar according to some predesignated criterion or criteria, while observations drawn from different clusters are dissimilar

- Different clustering techniques make different assumptions on the structure of the data, often defined by some similarity metric (similarity between members of the same cluster) and separation between different clusters.

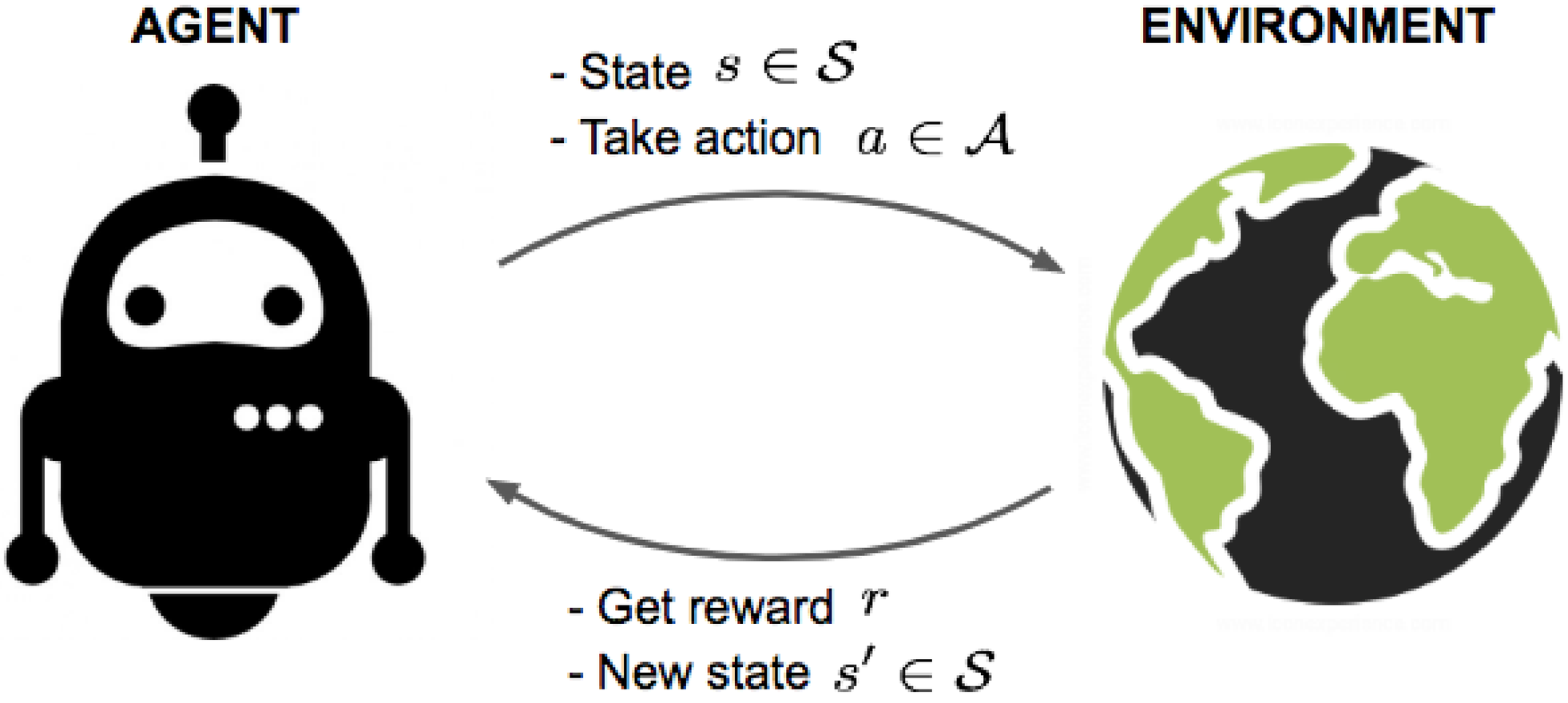
- Different types of clustering methodologies include centroid-based clustering, distribution-based clustering and density-based clustering

- **K-means** clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster

- https://en.wikipedia.org/wiki/Cluster_analysis



Reinforcement Learning

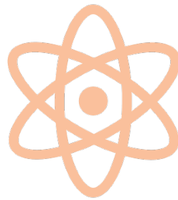


Reinforcement Learning Algorithms

- Reinforcement learning is concerned with how an *agent* ought to take *actions* in an *environment* so as to maximize some notion of long-term *reward* and *penalty*

- The basic reinforcement learning model consists of:
  - a set of environment states S;
  - a set of actions A;
  - rules of transitioning between states;
  - rules that determine the reward of a transition; and
  - rules that describe what the agent observes.

- Different approaches include:
  - Temporal difference learning
  - Q-learning
  - Learning Automata



3 Types of Reinforcement Learning

Model-based

- Learn the model of the world, then plan using the model
- Update model often
- Re-plan often

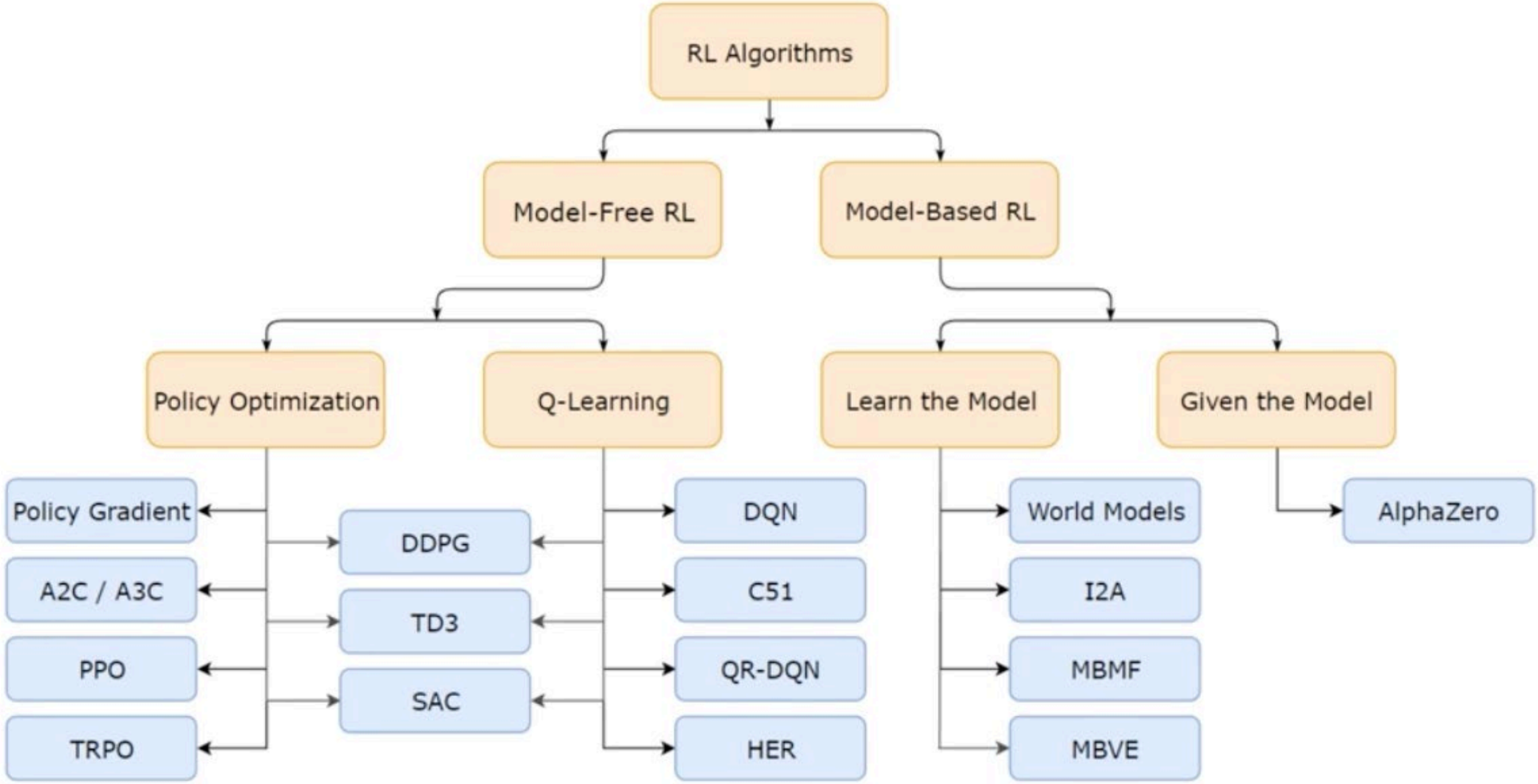
Value-based

- Learn the state or state-action value
- Act by choosing best action in state
- Exploration is a necessary add-on

Policy-based

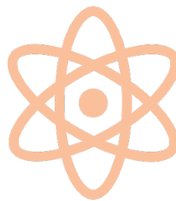
- Learn the stochastic policy function that maps state to action
- Act by sampling policy
- Exploration is baked in

Taxonomy of RL Methods

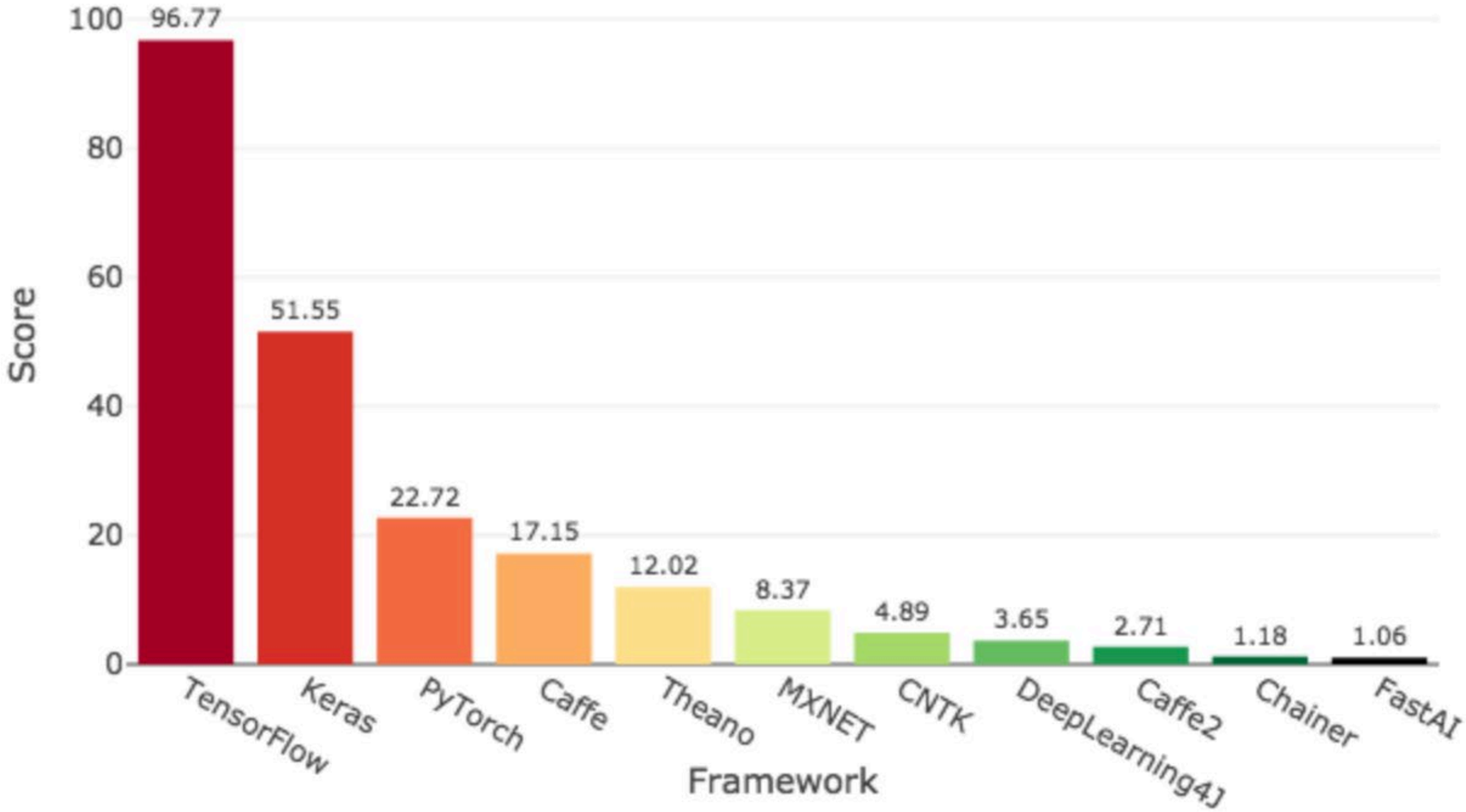


Link: https://spinningup.openai.com

Use (mostly) Python Libraries

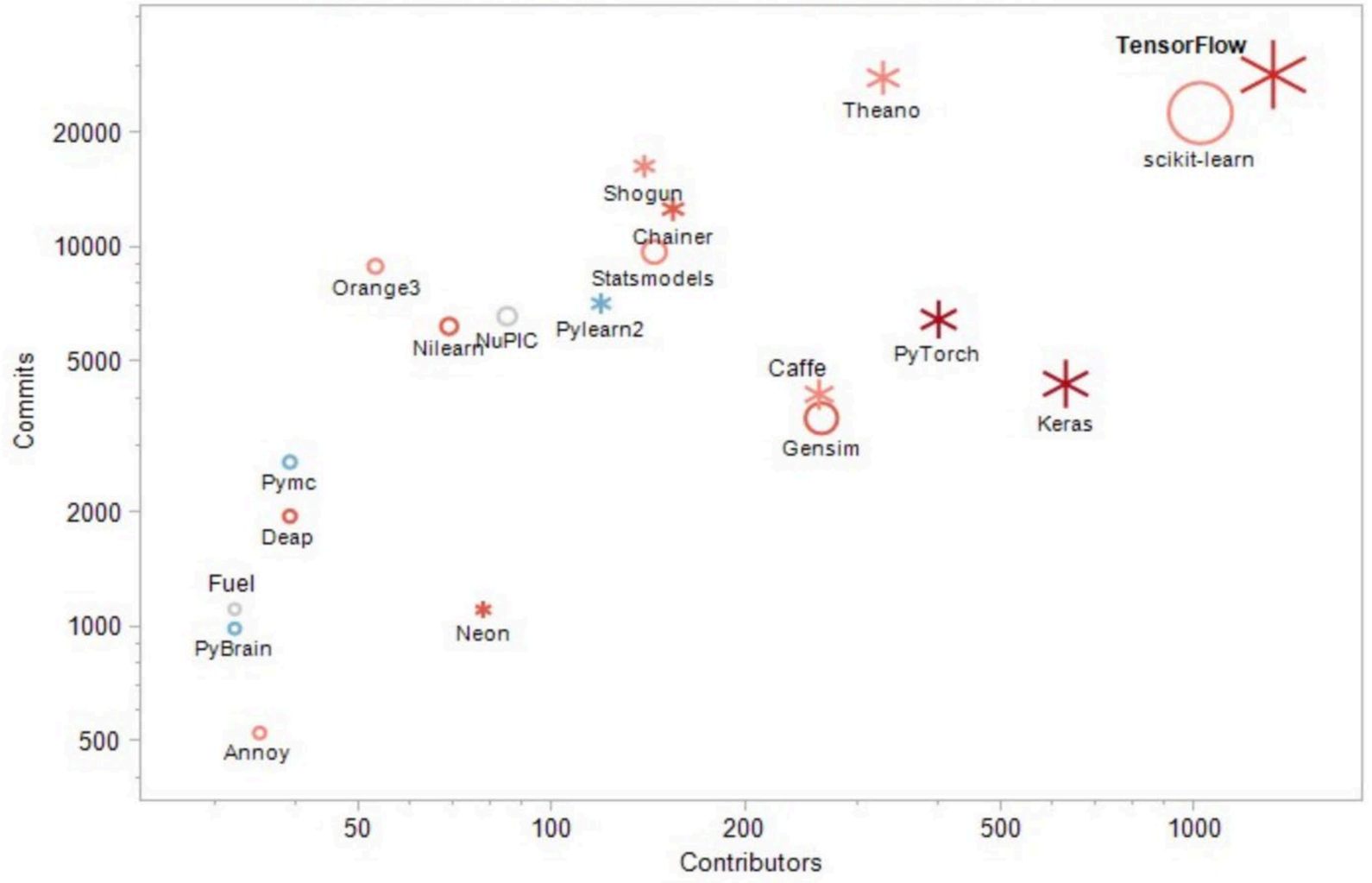


ML Frameworks - Popularity



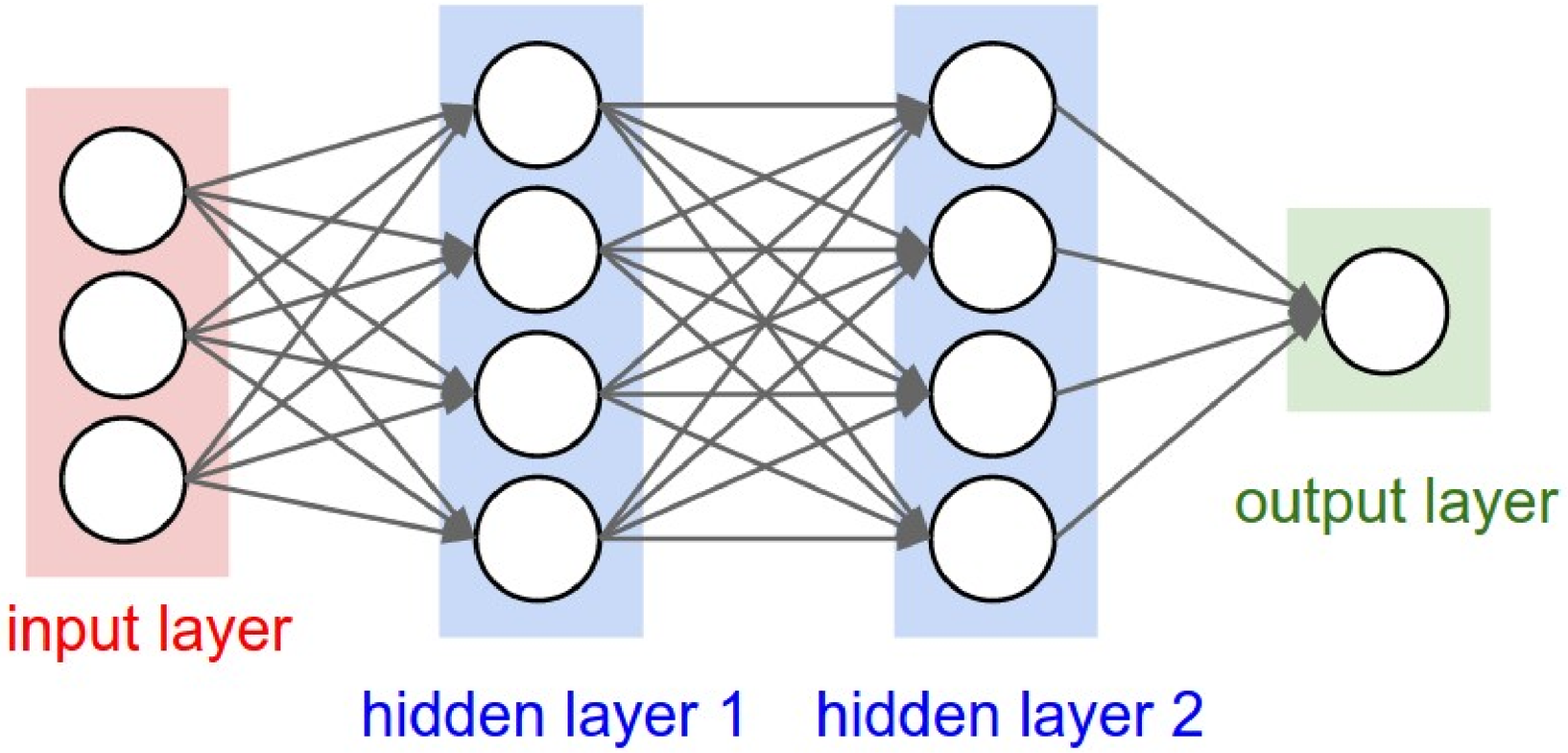
Towards Data Science, Sept 20, 2018 https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a

Top 20 Python AI and Machine Learning projects on Github



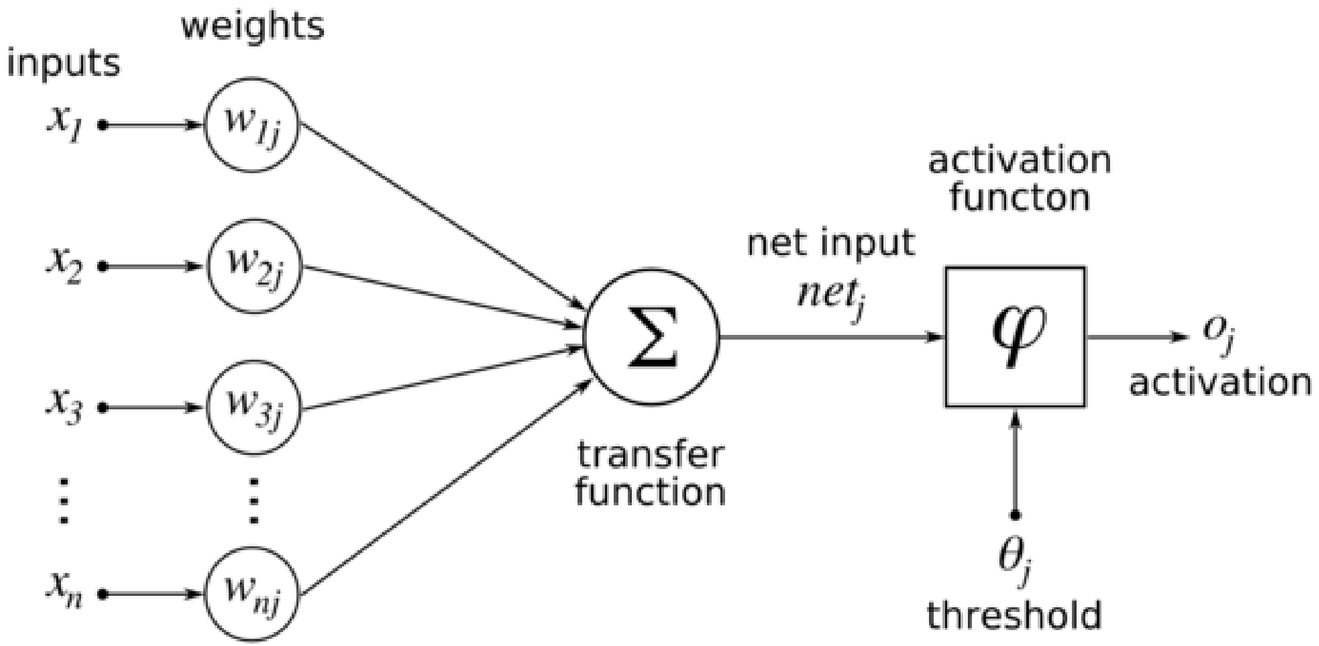
KD Nuggets, Feb 2018, https://www.kdnuggets.com/2018/02/top-20-python-ai-machine-learning-open-source-projects.html

Review: Neural Networks



Neural Network Nodes

X

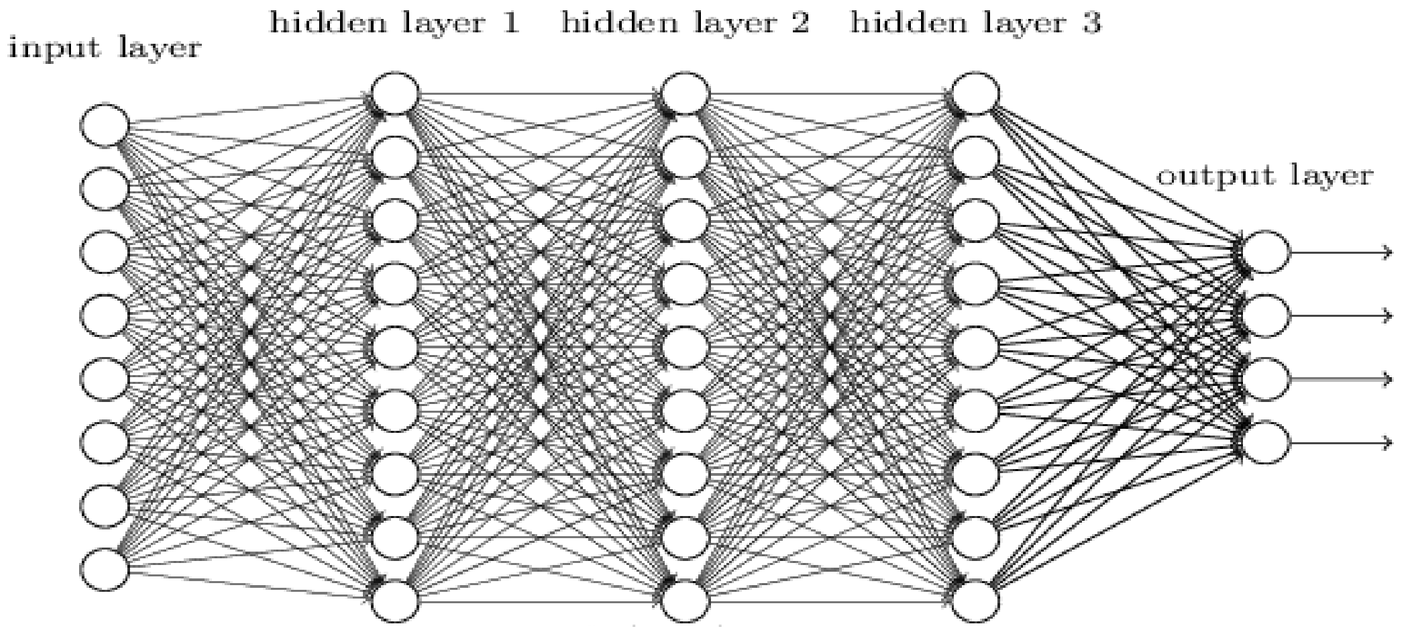


X'

$$X' = W.X + b$$

Multilayer Perceptron (MLP)

- A MLP is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs (Rosenblatt, 1958).
- An MLP consists of multiple layers of nodes with each layer fully connected to the next one.
- Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function.
- MLP utilizes a supervised learning technique called backpropagation for **training the network**.



Layers and Nodes

Layers

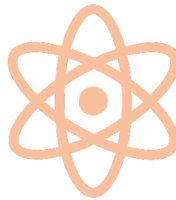
- A neural network is made up of layers of connected nodes **(like neurons and synapses).**
- There is an *input* (visible) layer, an *output* (classification) layer, and several *hidden* layers.
- A typical NN may have between 10 and 30 layers, but sometimes many more.
- Every layer of a deep learning network requires four elements: the input (vector), the weights (matrix), a bias and the transform (activation function).

Nodes

- Nodes in a neural network represent the places where calculations are done.
- At each node, the input values x_i are multiplied by a weight w_i, summed, added to a bias b, and fed into an activation function.
- A decision is then made whether to transmit the resultant value depending on if it exceeds a certain threshold value or not.
- For example, images from the MNIST data set have 784 pixels, so neural nets processing them must have 784 input nodes, one per pixel.

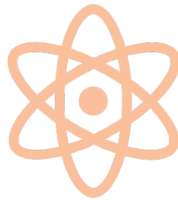
Weights

- Weights are coefficients that amplify or mute the input signal coming into each node. They take a value between 0 and 1.
- They are assigned initial values which can be random or chosen based upon some insight **(initialization).**
- A NN can be represented by its **weight matrix** along with its activation functions.
- Weights are like the synaptic connections in the brain – the higher the weight, the stronger the connection.
- The weights change as the network is trained on new data.



Activation Function

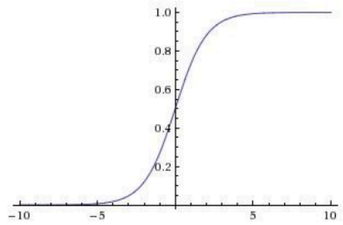
- One of a set of functions that determine the threshold at each node above which a signal is passed through the node, and below which it is blocked.
- Activation functions normalize input from a previous layer to a value between -1 and 1 to ensure an output layer probability of between 0 and 1.
- Functions that achieve this include the logistic, sigmoid, tanh and ReLU functions.
- In neural network simulations, the **softmax function** is often implemented at the final layer of a network used for classification.
- Activation Function https://en.wikipedia.org/wiki/Activation_function
- Softmax https://en.wikipedia.org/wiki/Softmax_function



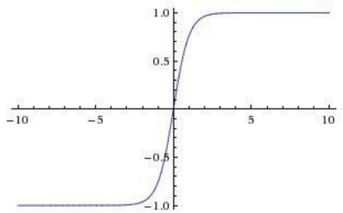
Activation Functions

Sigmoid

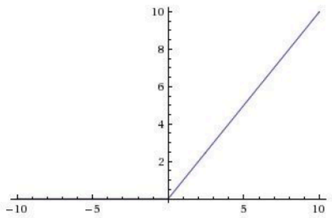
$$\sigma(x) = 1/(1+e^{-x})$$



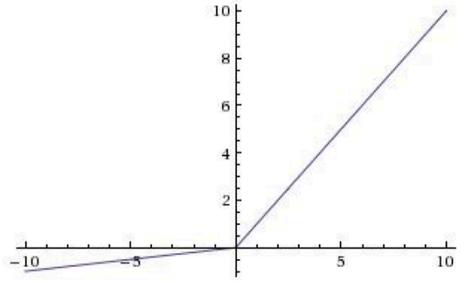
tanh tanh(x)



ReLU max(0,x)



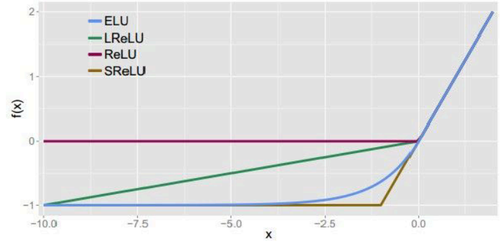
Leaky ReLU
max(0.1x, x)



Maxout $\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

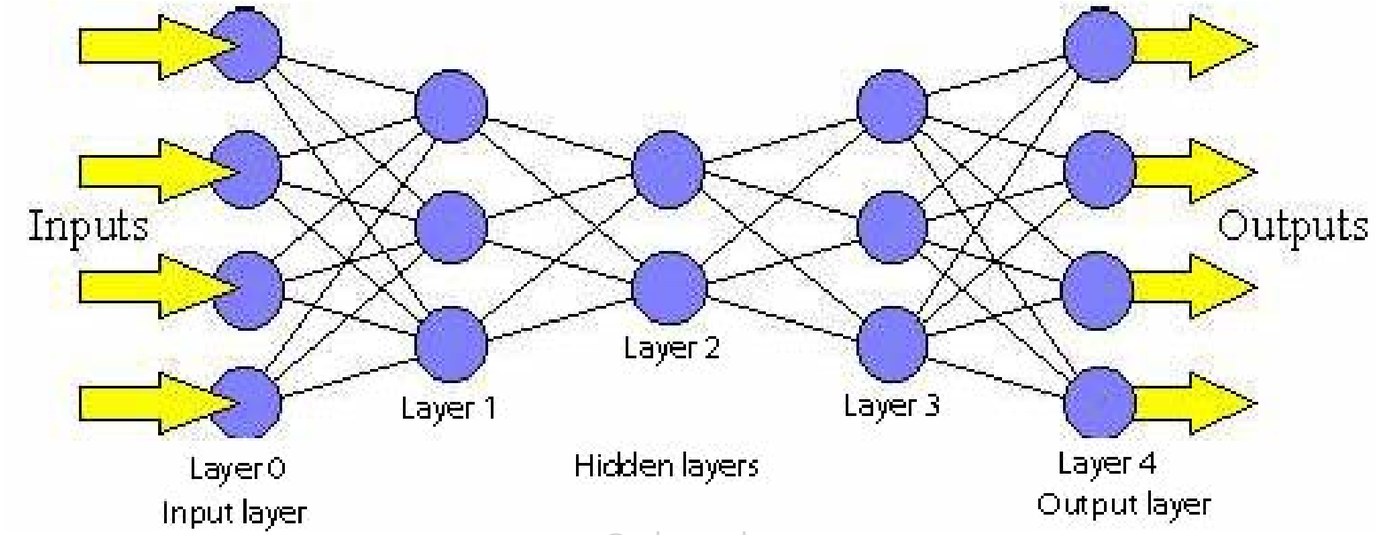
$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha\,(\exp(x) - 1) & \text{if } x \leq 0 \end{cases}$$



Source : http://cs231n.stanford.edu/slides/

Feed Forward Networks

- The feedforward neural network was the first and simplest type of artificial neural network devised.
- In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes.
- There are no cycles or feedback loops in the network.



Optimization and Overfitting

Optimization

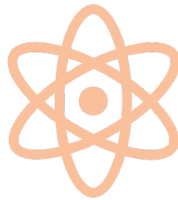
- Refers to the manner by which a neural net minimizes error as it adjusts its coefficients (weights) step by step.

- L-BFGS is one such algorithm.

Overfitting

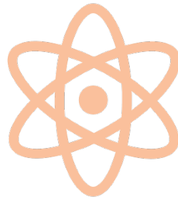
- Overfitting is where too many parameters are used in the model constructed to fit the data which leads to poor predictive power of the model.

Regularisation, cross-validation and dropout are all methods used to address overfitting.



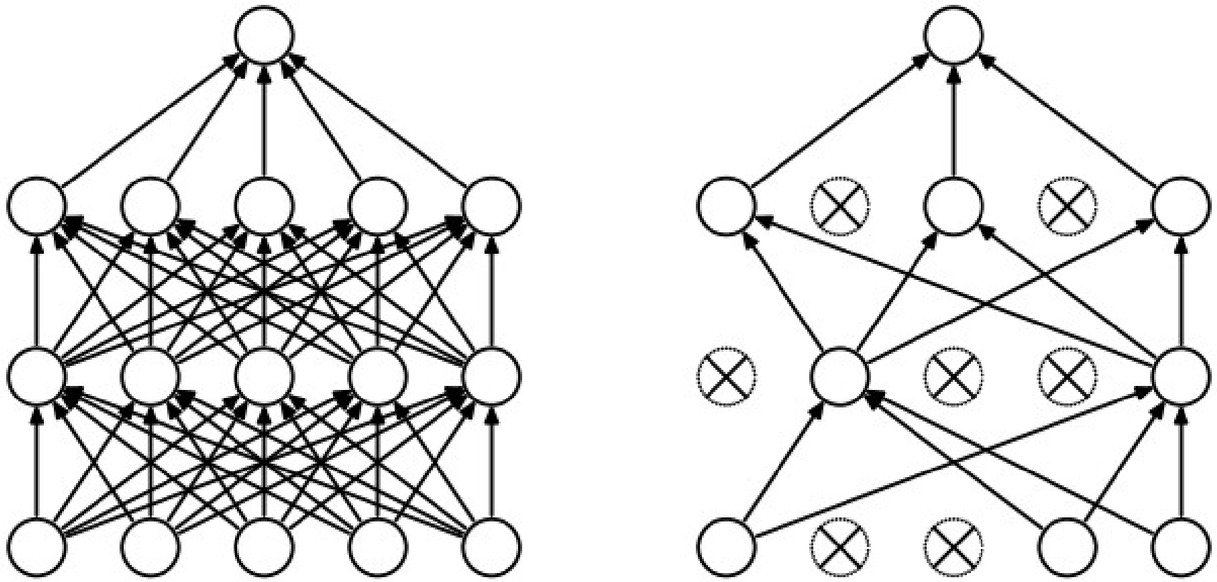
Overfitting Prevention Methods

- **Regularisation** refers to a process of introducing additional information in order to prevent overfitting.
- It penalizes models with extreme parameter values by introducing a factor which weights the penalty against more complex models with an increasing variance in the data errors.
- **Cross-validation** combines sampling sets to correct for overfitting and derive a more accurate estimate of model prediction performance.
- One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set).
- To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.
- **Dropout:** By avoiding training all nodes on all the training data, dropout decreases overfitting in neural nets. The method also significantly improves the speed of training.
- At each training stage, individual nodes are either "dropped out" of the net with probability 1-p or kept with probability p, so that a reduced network is left.



Dropout

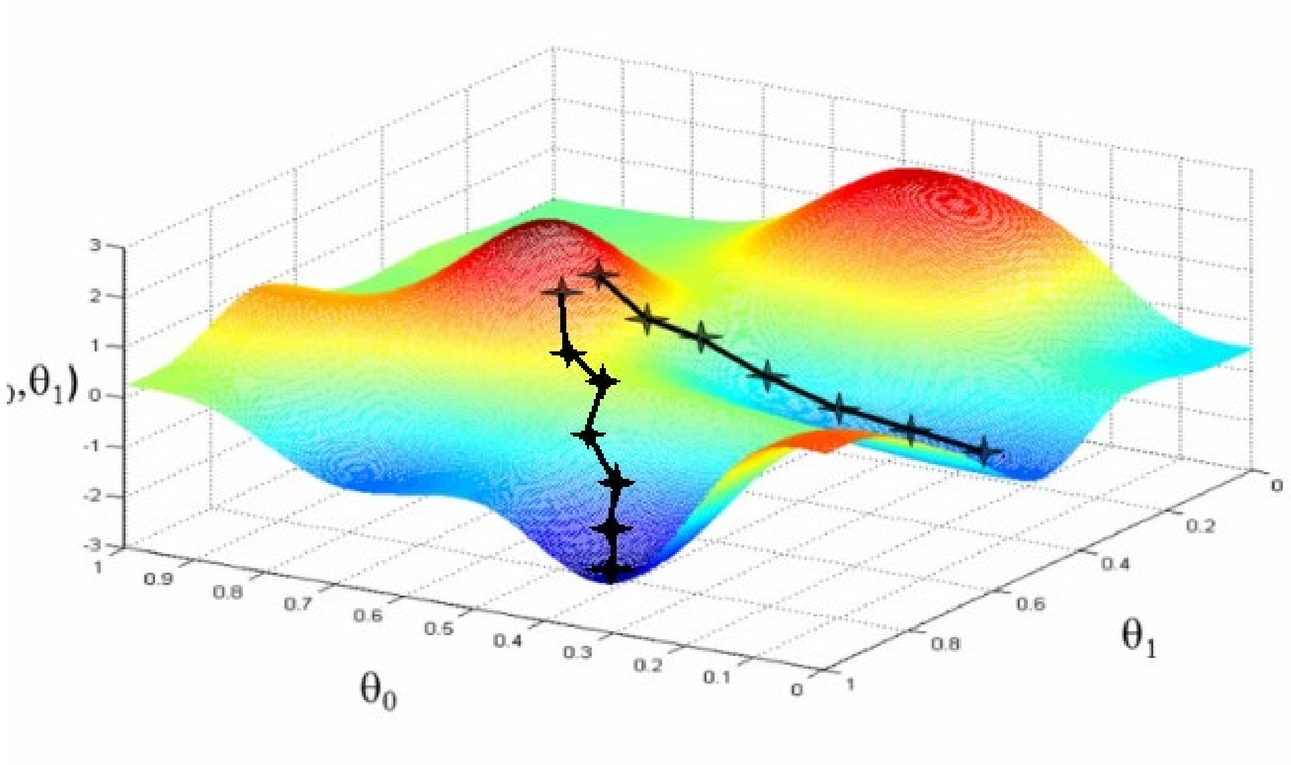
- To reduce over fitting, we will apply dropout before the readout layer
- Dropout is an extremely effective, simple and recently introduced regularization technique by Srivastava et al. in "Dropout: A Simple Way to Prevent Neural Networks from Overfitting" that complements the other methods (L1, L2, maxnorm).



(a) Standard Neural Net

(b) After applying dropout.

Stochastic Gradient Descent



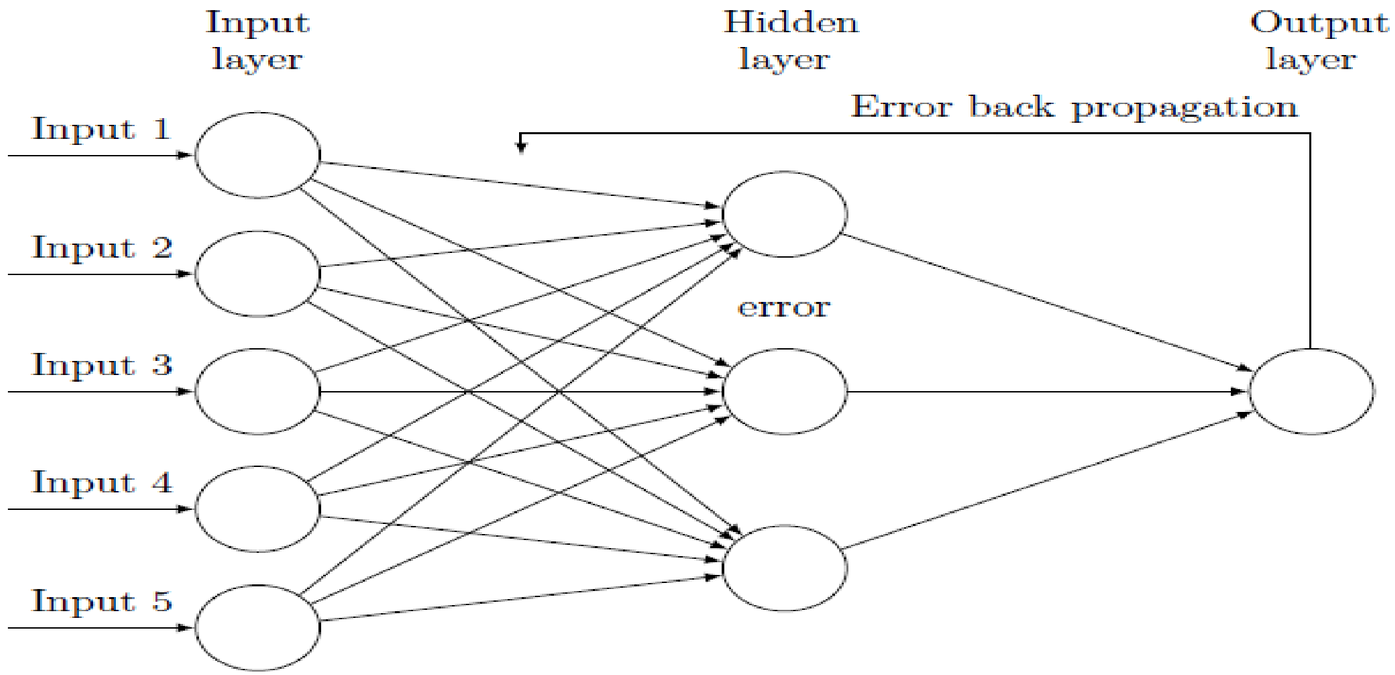
- Stochastic gradient descent is a popular algorithm for training a wide range of models in machine learning, including support vector machines and logistic regression.
- When combined with the backpropagation algorithm, it is the de facto standard algorithm for training artificial neural networks.

Back Propagation

- Backpropagation is a common method of **training artificial neural networks** used in conjunction with an optimization method such as **gradient descent.**

- The method calculates the gradient of a loss function with respect to all the weights in the network.

- The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to **minimize the loss function.**



Restricted Boltzmann Machines & Autoencoders

- **Restricted Boltzmann Machines (RBM)**

- Invented by Geoff Hinton (1999) at the University of Toronto, RBMs are shallow, **two-layer neural nets** consisting of an input layer and a hidden layer.

- They constitute the building blocks of *deep neural networks* - a deep learning network is simply many restricted Boltzmann machines **stacked on top of one another.**

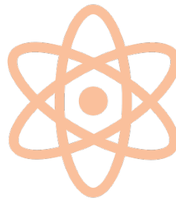
- Nodes are connected to each other across layers, but have the restriction that **no two nodes of the same layer are linked.**

- **Autoencoder**

- An autoencoder is an ANN used for learning efficient codings.

- The aim of an autoencoder is to learn a compressed, distributed representation (encoding) for a set of data, typically for the purpose of dimensionality reduction.

- In an autoencoder, the output layer has equally many nodes as the input layer, and instead of training it to predict some target value *y* given inputs x, an autoencoder is trained to **reconstruct its own inputs x.**



Max Pooling and Cost Functions

Max Pooling

- In CNN's, max-pooling partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum value.

Cost Function

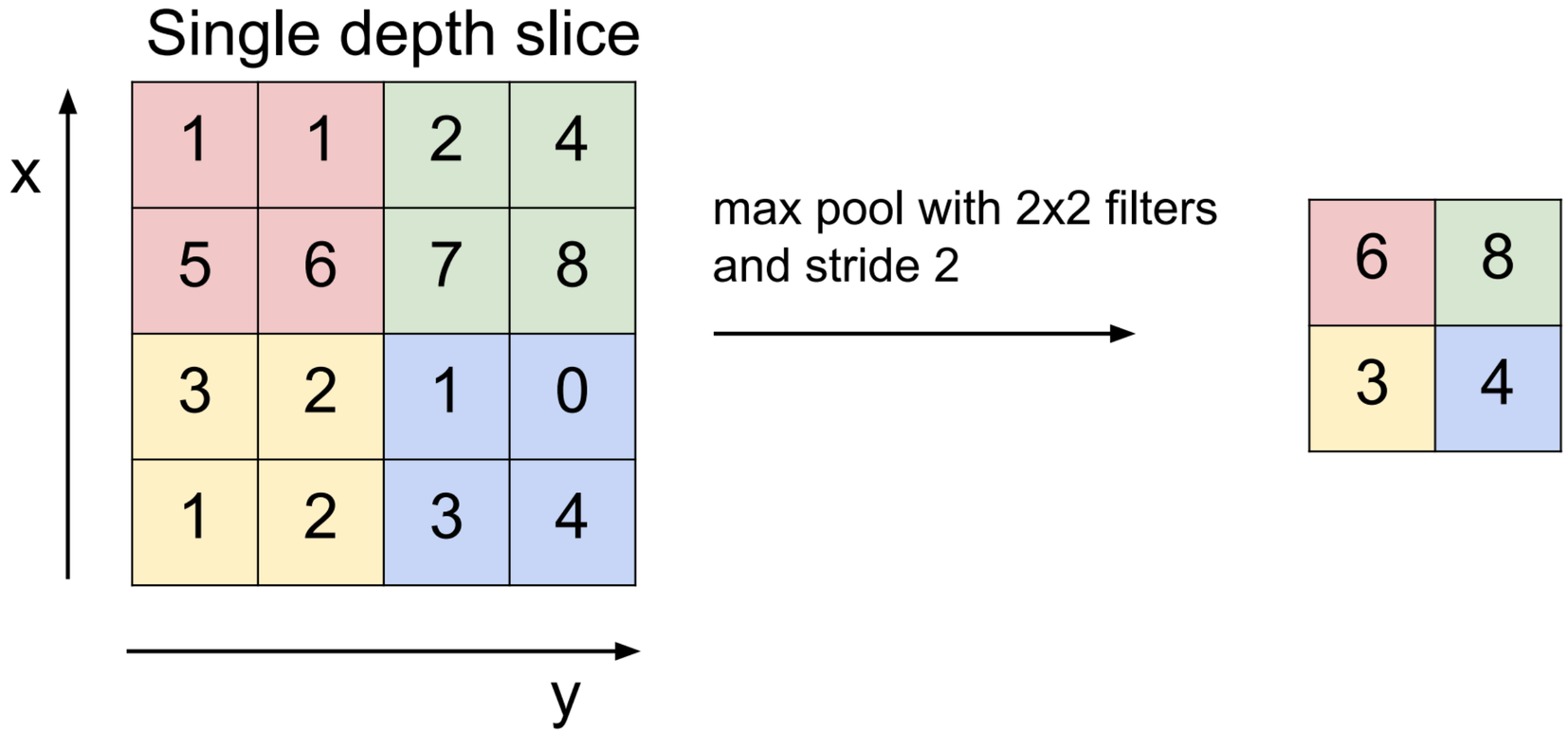
- A cost function or loss function is a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event.

- The opposite of a loss function is called a reward function, or utility function.

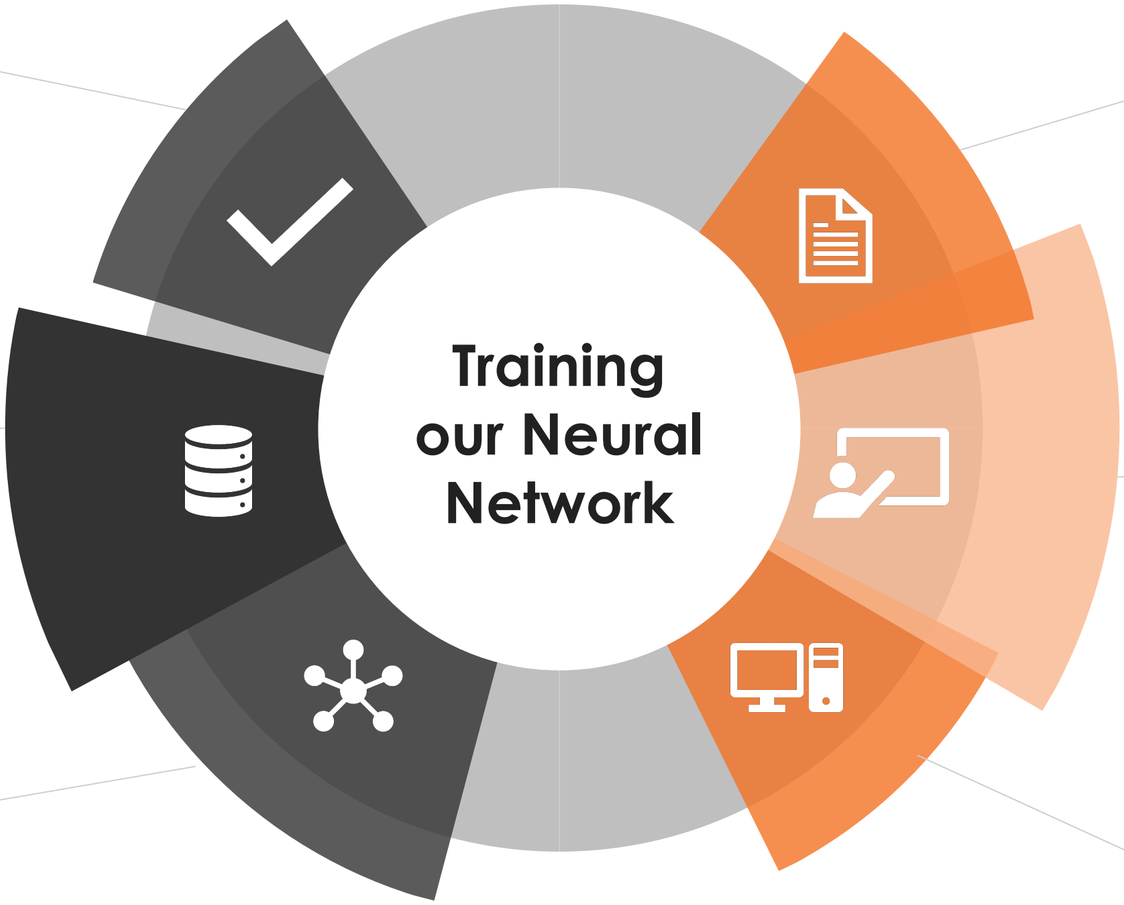
- The difference (delta) between the output and the ground truth is a cost

Pooling Layer

MAX POOLING



After we have identified our data set, there are generally 5 steps in training a neural network:



Load and normalize data

Choose the neural network architecture

Choose the loss function

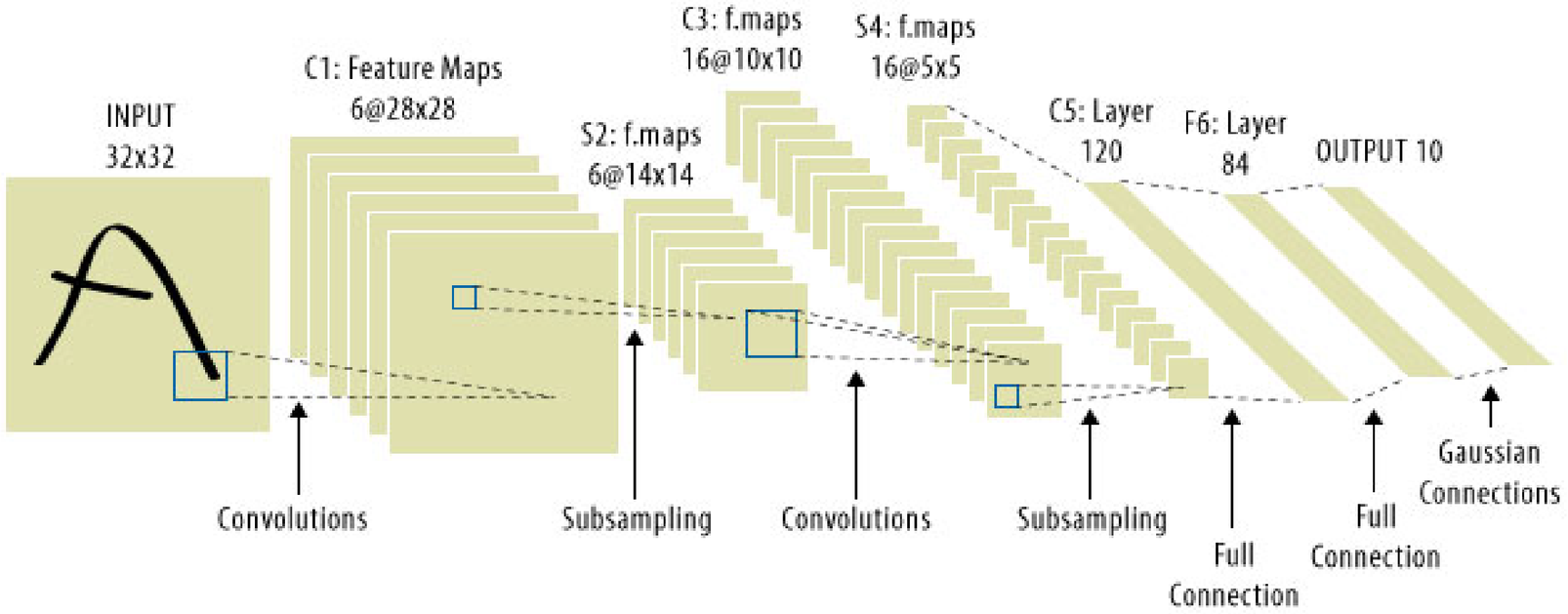
Train the network on training data

Test the network on test data (inference).

Computer Vision

- Computer Vision is the ability of computers to see
- CV seeks to automate tasks that the human visual system can do
- Includes digital images or videos
- The raw data is light waves (analog) or pixels (digital). Analysis usually involves vectorising pixels
- Image understanding can be seen as using models constructed with the aid of geometry, physics, statistics, and learning theory
- Computer vision is concerned with the theory behind artificial systems that extract information from images then seeks to apply these theories to the construction of computer vision systems
- It involves the fields of pattern recognition and image and signal processing

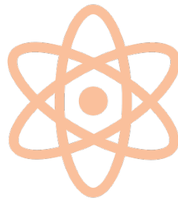
Convolutional Neural Networks



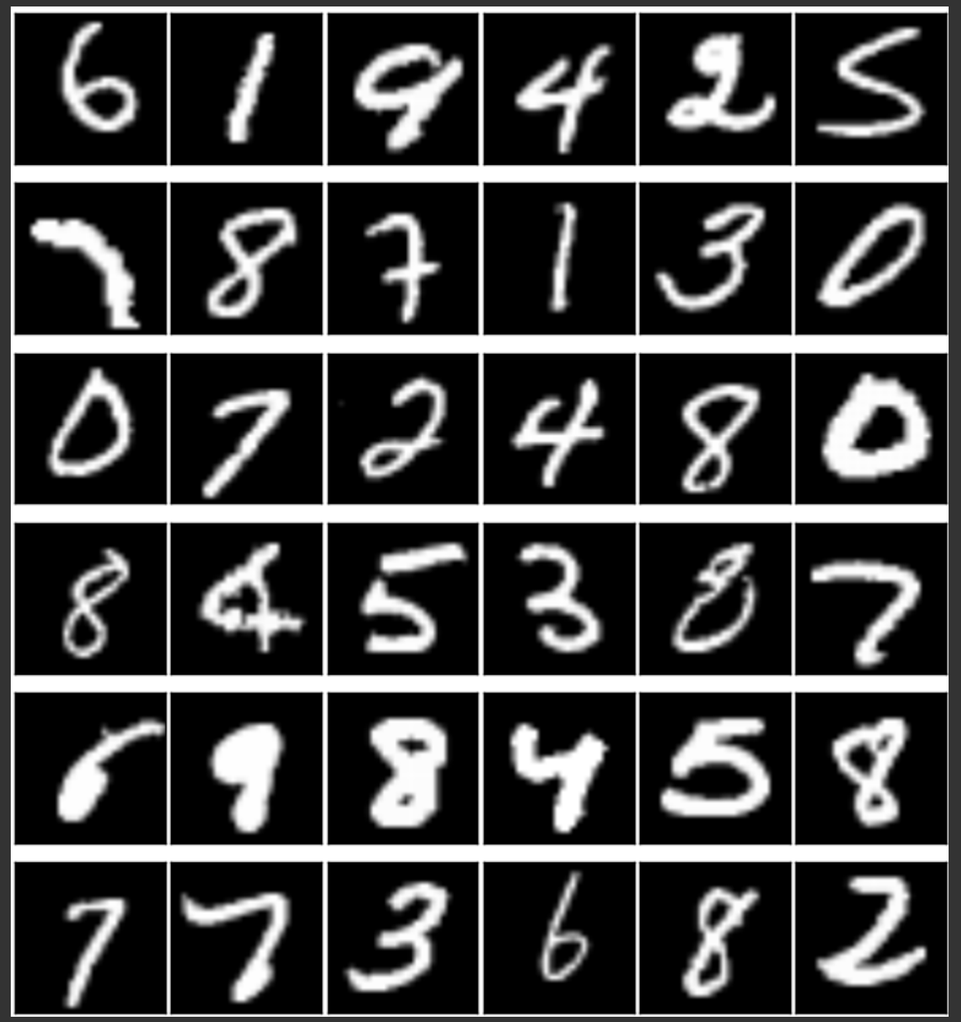
- First developed in 1970's.
- Widely used for image recognition and classification.
- Inspired by biological processes, CNN's are a type of feed-forward ANN.
- The individual neurons are tiled in such a way that they respond to overlapping regions in the visual field.

Image Data Sets

- MNIST http://yann.lecun.com/exdb/mnist/
- CIFAR-10 & 100 http://www.cs.toronto.edu/~kriz/cifar.html
- ImageNet http://www.image-net.org/
- SVHN http://ufldl.stanford.edu/housenumbers/
- Labeled Faces in the Wild http://vis-www.cs.umass.edu/lfw/index.html
- Caltech 101 http://www.vision.caltech.edu/Image_Datasets/Caltech101
- Various http://www.face-rec.org/databases/



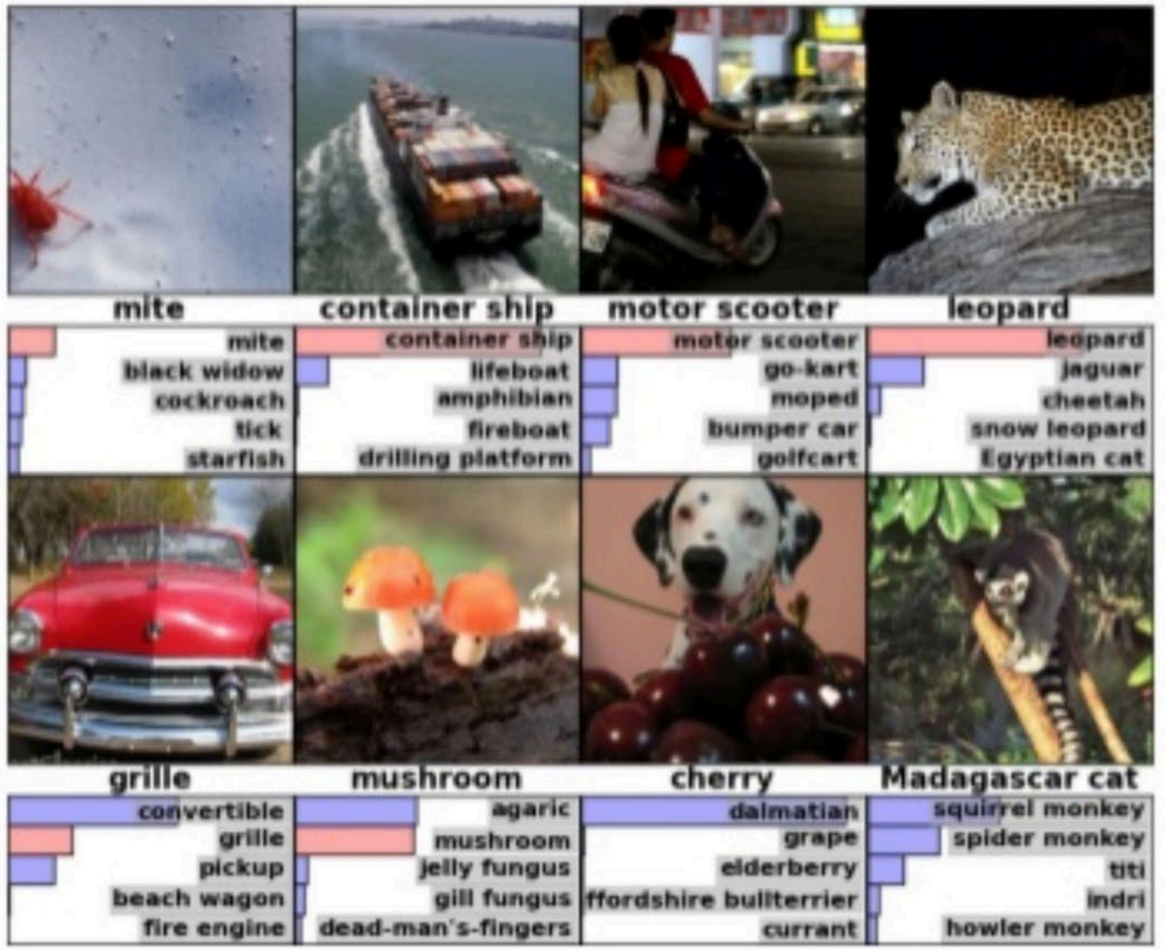
MNIST Dataset



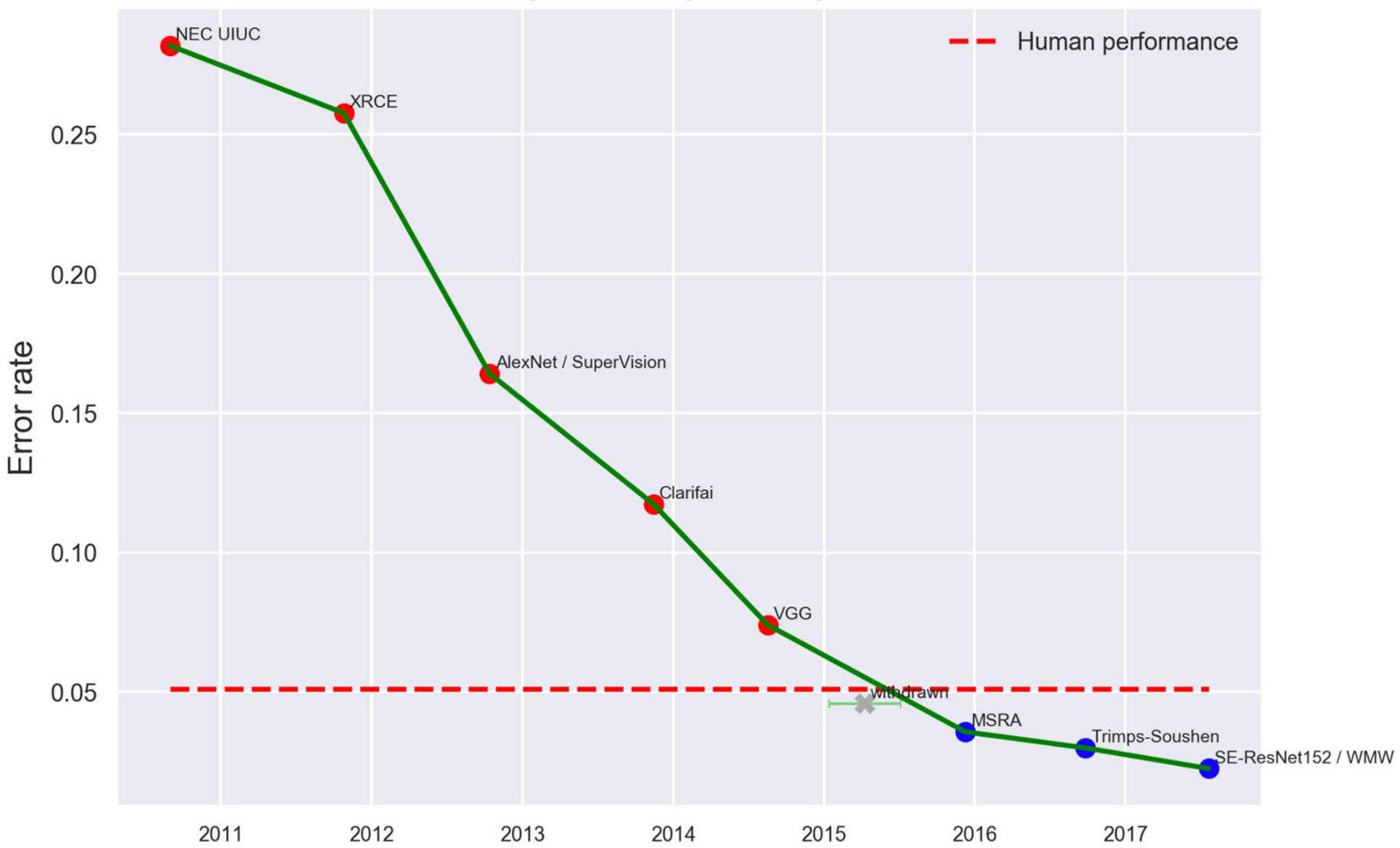
MNIST Samples

IMAGENET

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



Imagenet Image Recognition



Data Augmentation

Crop:

Flip:

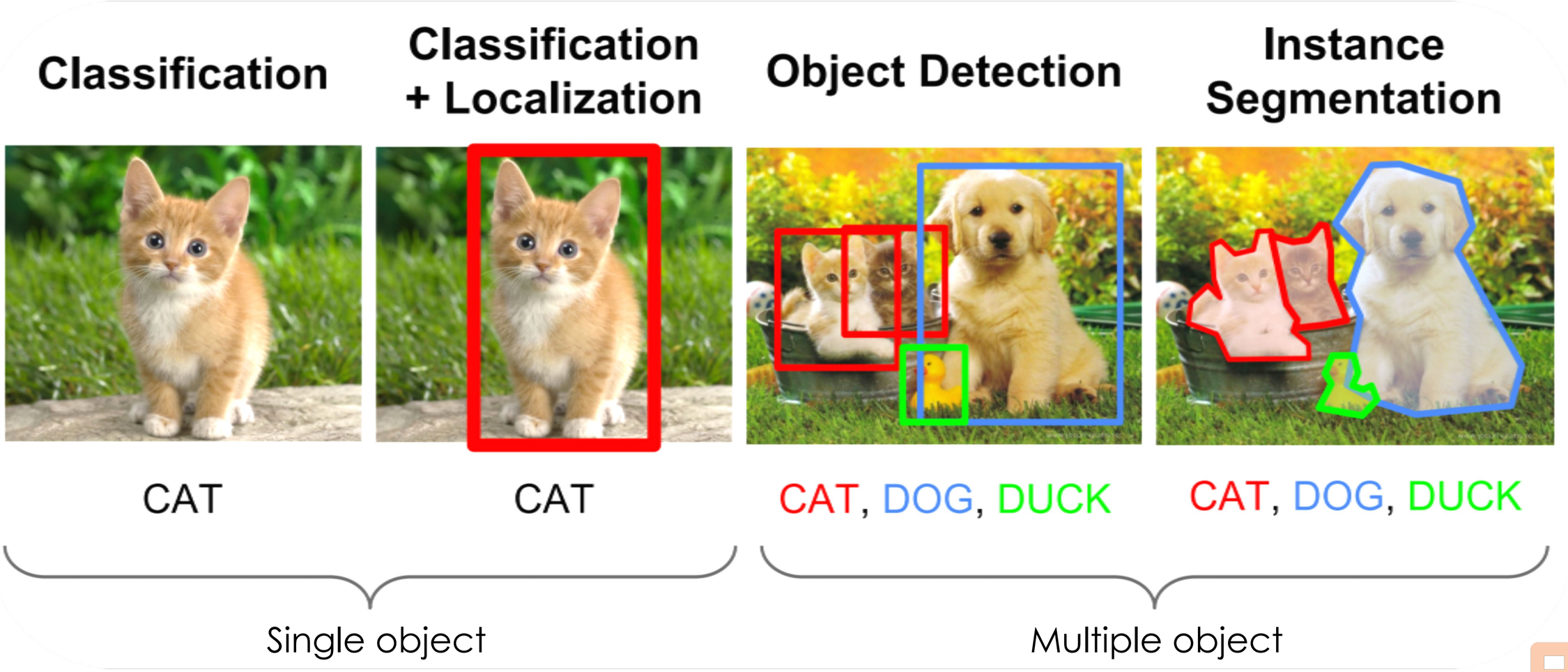
Scale:

Rotate:

Translation:

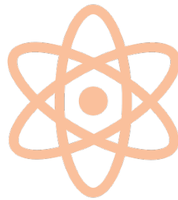
Noise:

Computer Vision Tasks



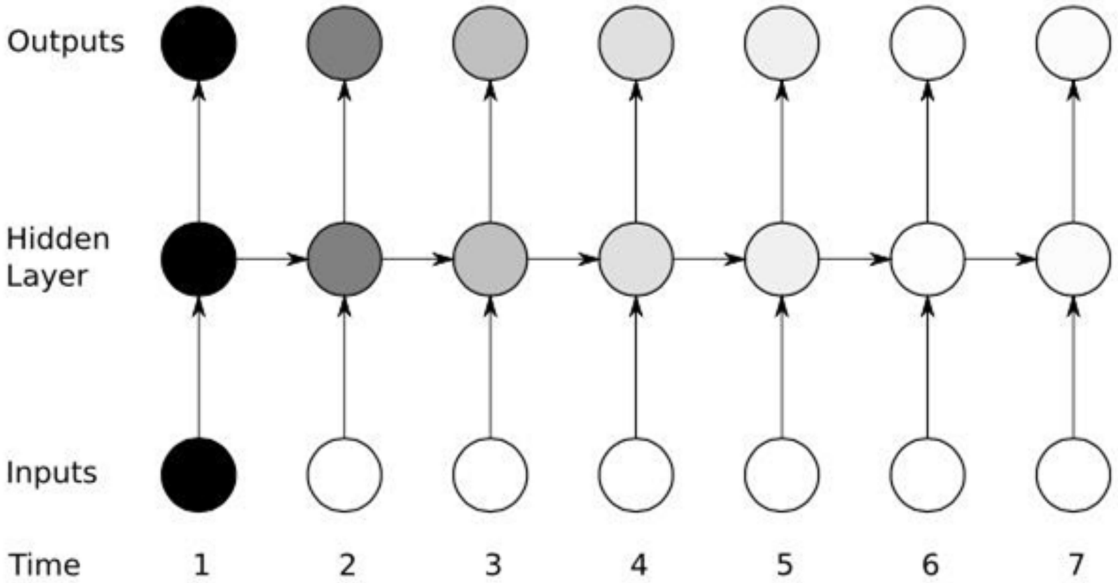
Natural Language Processing

- Natural Language Processing is concerned with how computers process spoken or written natural language
- Recognizing words, their meaning, understand context and narrative
- Many challenges in NLP involve natural language understanding, i.e., enabling computers to derive meaning from human language input
- Others involve natural language generation



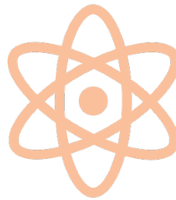
Recurrent Neural Networks

- First developed in 1970's.
- RNN's are neural networks that are used to predict the next element in a sequence or **time series**.
- This could be, for example, words in a sentence or letters in a word.
- Applications include **predicting or generating** music, stories, news, code, financial instrument pricing, text, speech, video, in fact the next element in any ***event stream*.**



LSTM

- **Long Short Term Memory (LSTM)**
- LSTM (Schmidhuber, 1997) is an RNN architecture that contains blocks that can remember a value for an **arbitrary length of time.**
- It solves the vanishing or exploding gradient problem when calculating back propagation.
- An LSTM network is **universal** in the sense that given enough network units it can compute anything a conventional computer can compute, provided it has the proper weight matrix.
- LSTM **outperforms** alternative RNNs and Hidden Markov Models and other sequence learning methods in numerous applications, e.g., in handwriting recognition, speech recognition and music composition.



Speech and Sound

Sound Analysis

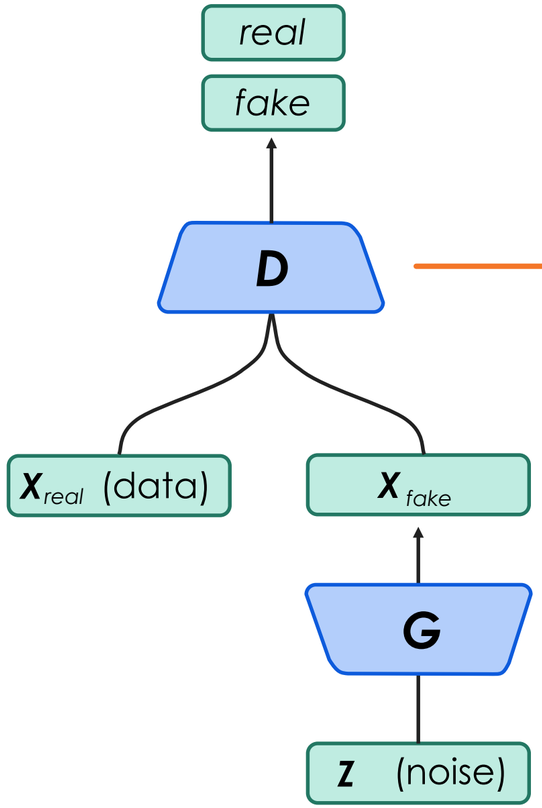
- Speech recognition, and in the broader sense audio signal processing, examines how machines and humans process audio signals
- It is concerned with the recognition and classification of sound (including noise) from any source - human, animal, environmental, speech, animate or inanimate
- Analysis of sound is performed upon the sound waves being incident on the receiver
- Segregation of foreground and background events and discrimination of various sounds and speakers constitute current areas of active research in speech recognition

GANs

GANs are used to generate things – images, videos, music, …

GAN - General Adversarial Networks

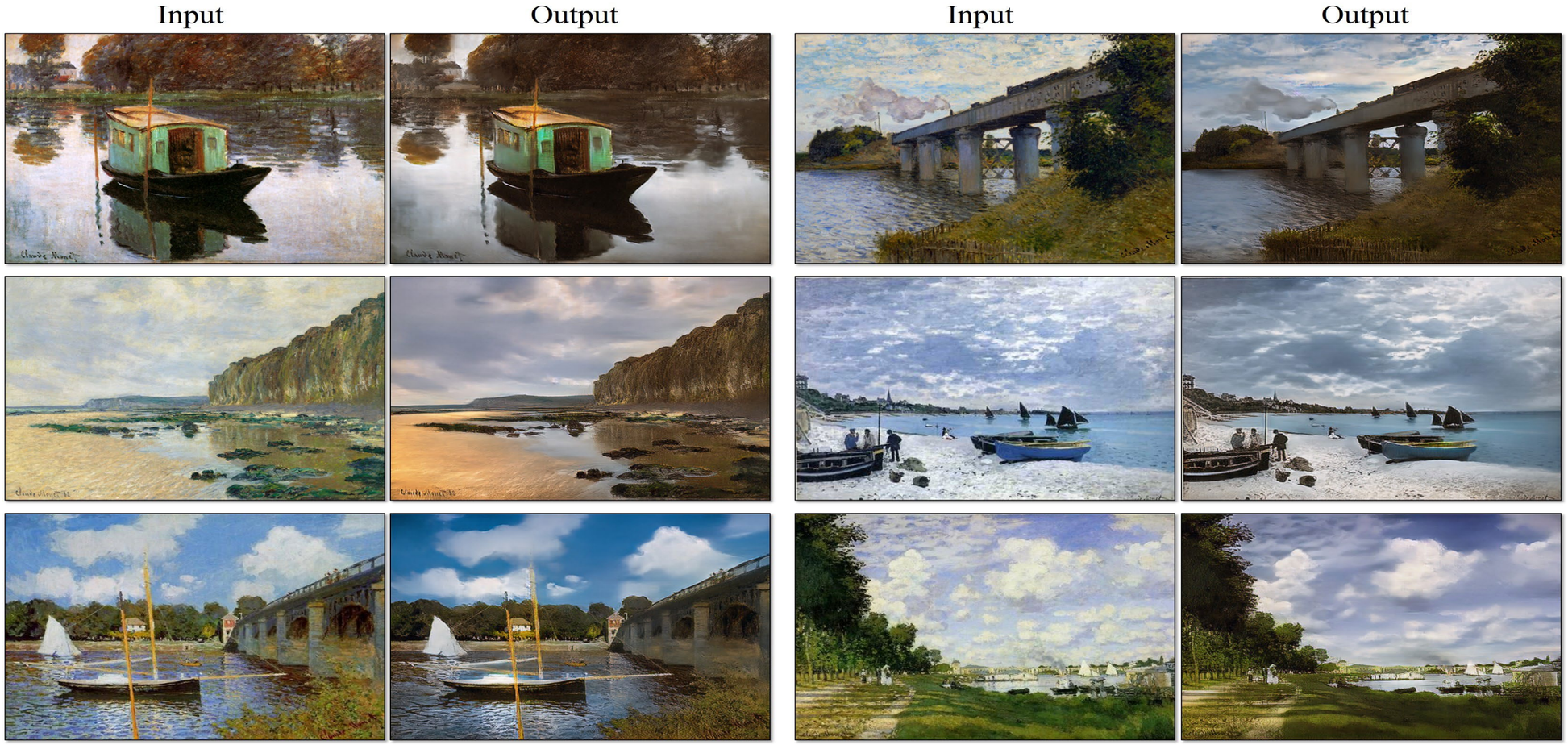
Generative Adversarial Networks (GANs) are a way to make a generative model by having two neural networks compete with each other.



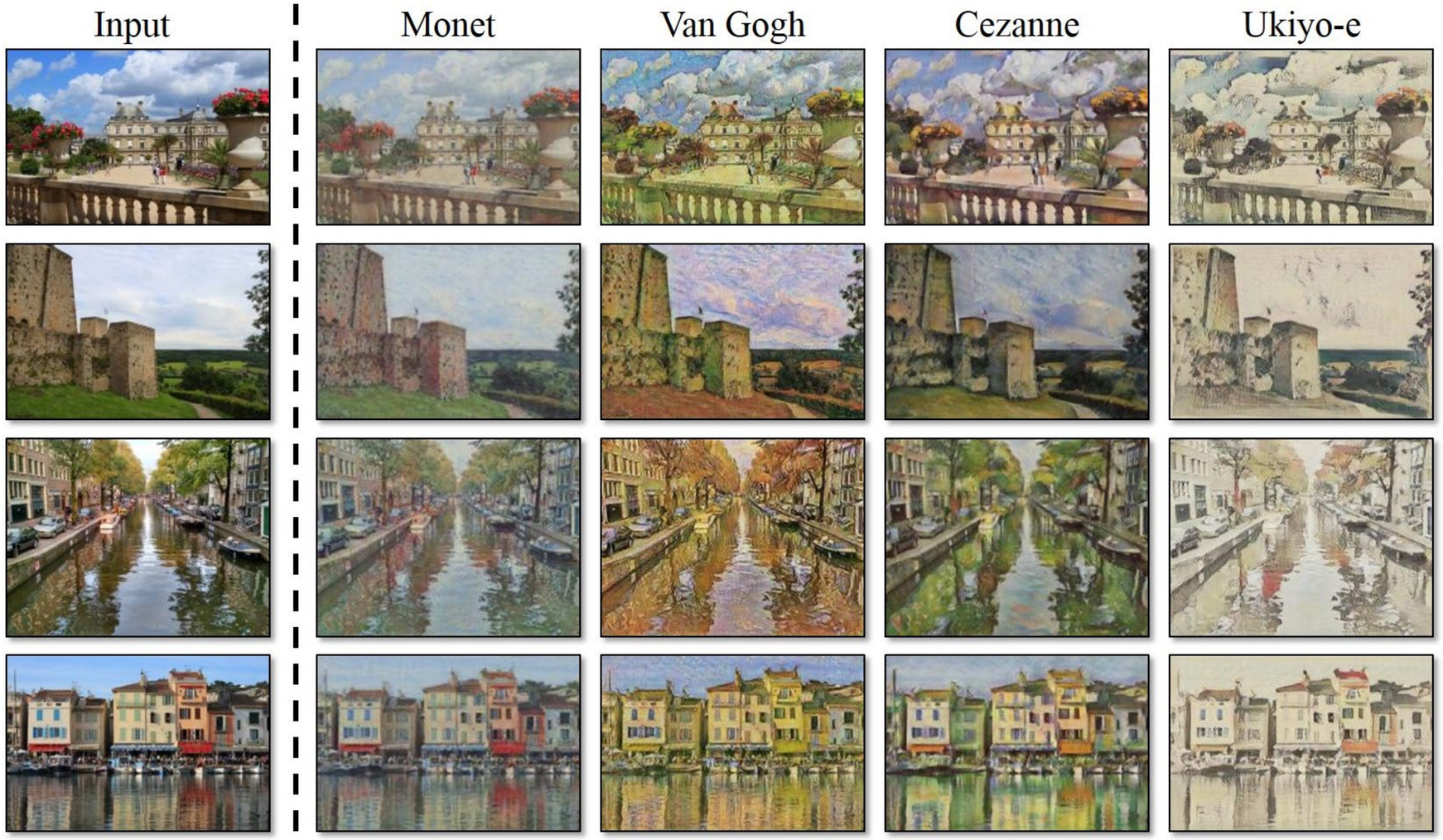
The **discriminator** tries to distinguish genuine data from forgeries created by the generator.

The **generator** turns random noise into imitations of the data, in an attempt to fool the discriminator.

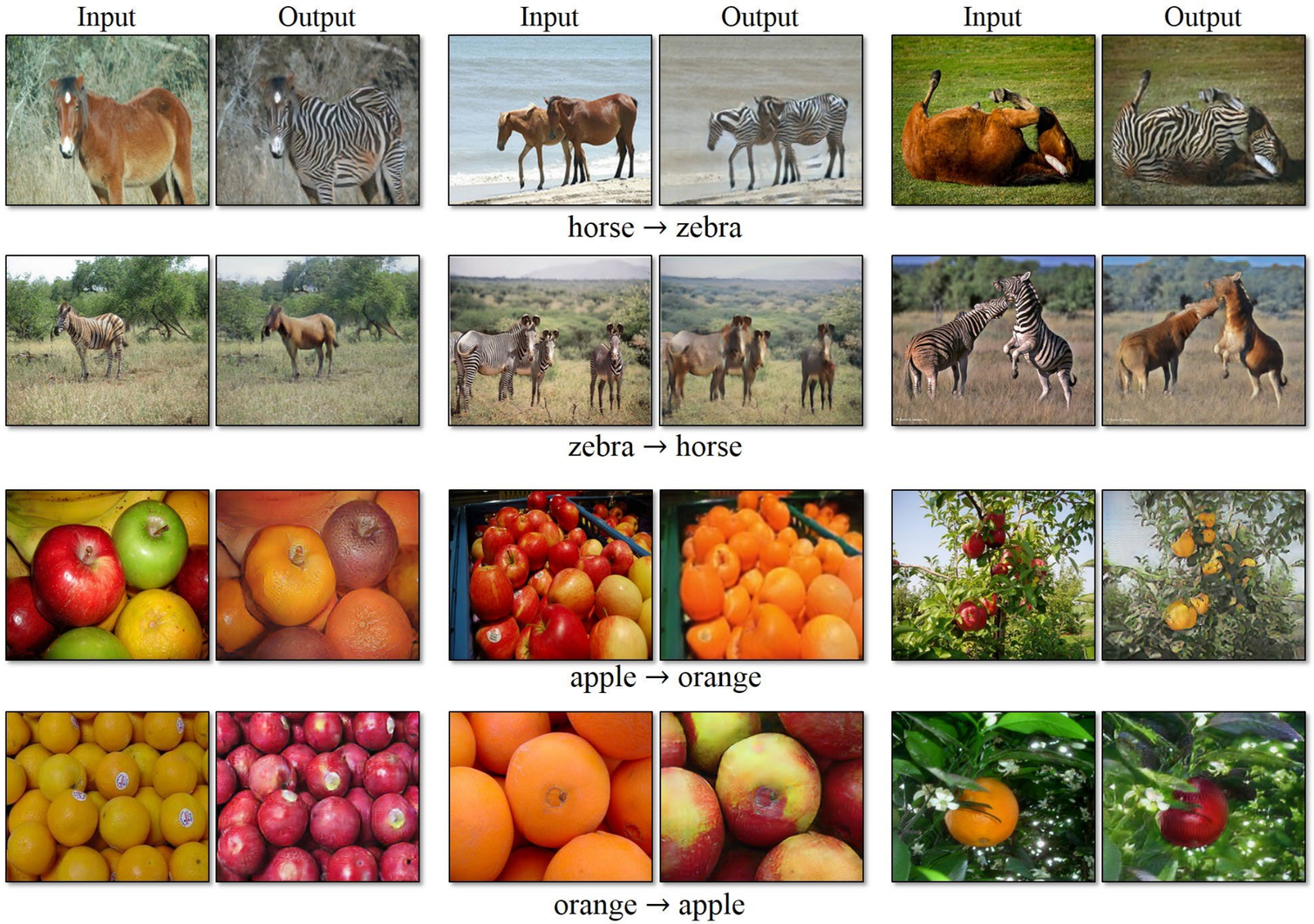
Paintings to Photos



Style Transfer

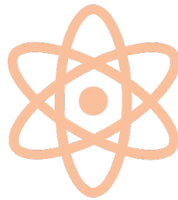


Object Transfiguration



GANS - References

- Generative Adversarial Nets in TensorFlow — https://wiseodd.github.io/techblog/2016/09/17/gan-tensorflow/
- Keeping up with the GANs — https://medium.com/nurture-ai/keeping-up-with-the-gans-66e89343b46
- The GAN Zoo — https://deephunt.in/the-gan-zoo-79597dc8c347
- Ian Goodfellow talk on GANs — https://www.youtube.com/watch?v=HN9NRhm9waY
- GAN code — https://github.com/carpedm20/DCGAN-tensorflow
- GAN applications — https://github.com/nashory/gans-awesome-applications
- GAN series — https://medium.com/@jonathan_hui/gan-gan-series-2d279f906e7b



TensorFlow

TensorFlow in one Slide

- **What is it:** Deep Learning Library *(and more)*
  - **Facts:** Open Source, Python, Google

- **Community:**
  - 117,000+ GitHub stars
  - TensorFlow.org: Blogs, Documentation, DevSummit, YouTube talks

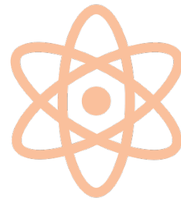
Ecosystem:

- Keras: high-level API
- TensorFlow.js: in the browser
- TensorFlow Lite: on the phone
- Colaboratory: in the cloud
- TPU: optimized hardware
- TensorBoard: visualization
- TensorFlow Hub: graph modules

Extras:

- Swift for TensorFlow
- TensorFlow Serving
- TensorFlow Extended (TFX)
- TensorFlow Probability
- Tensor2Tensor

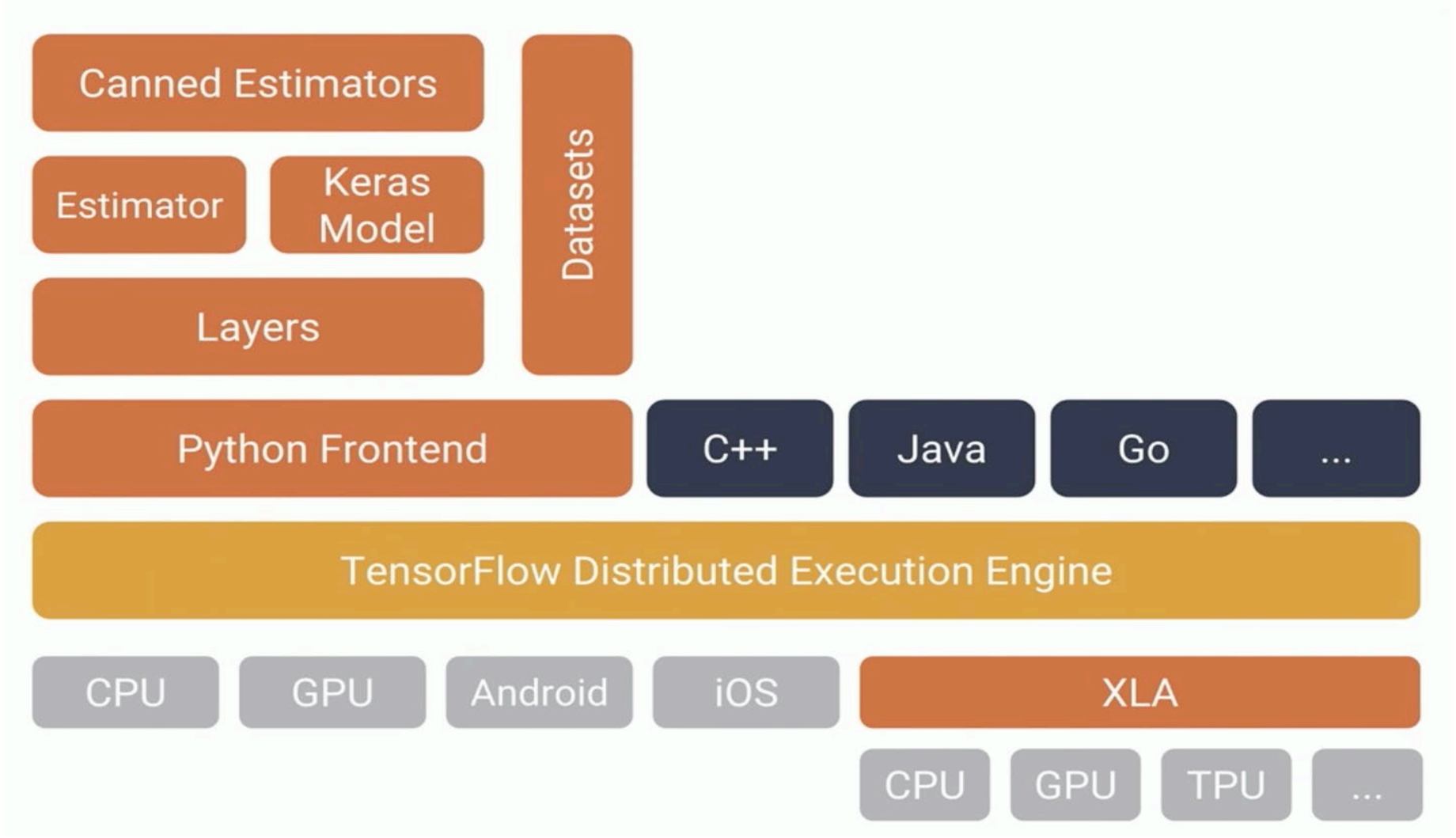
- **Alternatives:** PyTorch, MXNet, CNTK



Activity

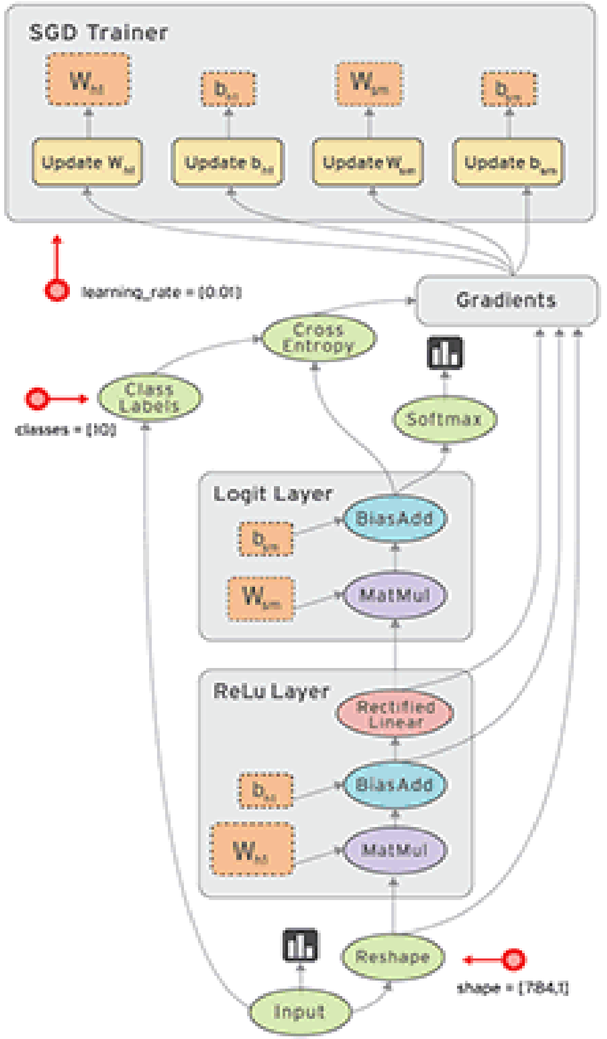
TensorFlow Website http://www.tensorflow.org

TensorFlow Architecture

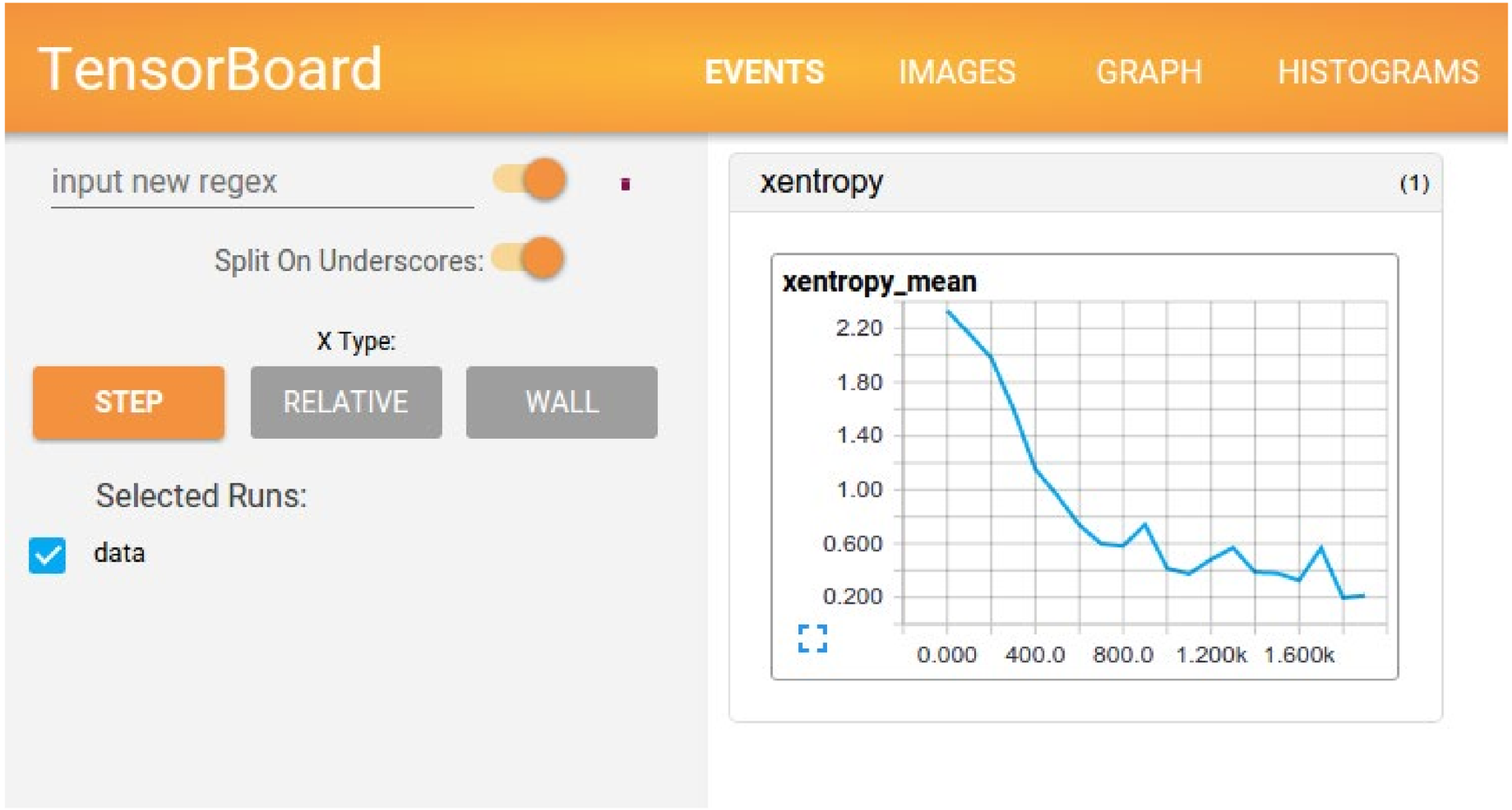


Data Flow Graph

- Data flow graphs describe mathematical computation with a directed graph of nodes & edges
- **Nodes** in the graph represent mathematical operations.
- **Edges** represent the multidimensional data arrays (tensors) communicated between them.
- **Edges** describe the input/output relationships between nodes.
- The flow of tensors through the graph is where **TensorFlow** gets its name.



TensorBoard : Visual Learning



TensorFlow Lite (Mobile)

TensorFlow Lite is for mobile and embedded devices

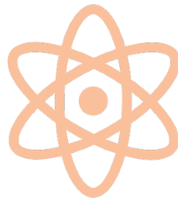
TensorFlow Lite is the official solution for running machine learning models on mobile and embedded devices. It enables on-device machine learning inference with low latency and a small binary size on Android, iOS, and other operating systems.

Many benefits

On-device ML inference is difficult because of the many constraints—TensorFlow Lite can solve these:

- **Performance**
  TF Lite is fast with no noticeable accuracy loss—see the metrics.
- **Portability**
  Android EI, iOS, and more specialized IoT devices.
- **Low latency**
  Optimized float- and fixed-point CPU kernels, op-fusing, and more.
- **Acceleration**
  Integration with GPU and internal/external accelerators.
- **Small model size**
  Controlled dependencies, quantization Z, and op registration.
- **Tooling**
  Conversion, compression, benchmarking, power-consumption,and more.

https://www.tensorflow.org/lite



Production Pipeline

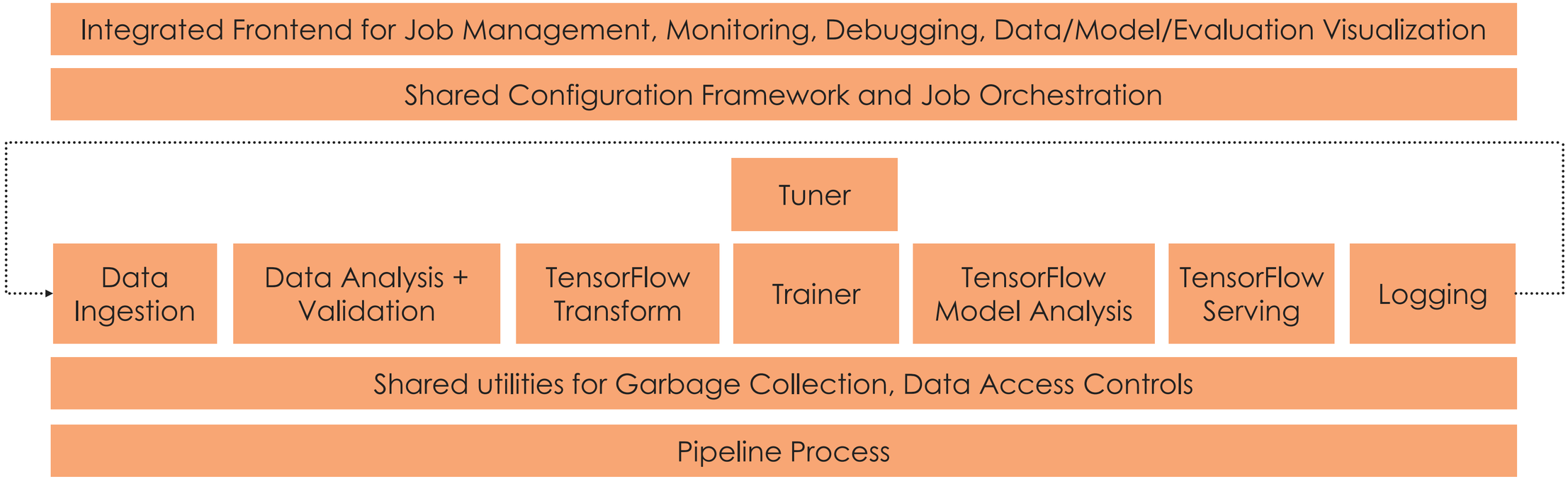


Figure 1: High-level component overview of a machine learning platform.

Demo 1

Get Started with TensorFlow

5 mins

https://www.tensorflow.org/tutorials/

Demo 2

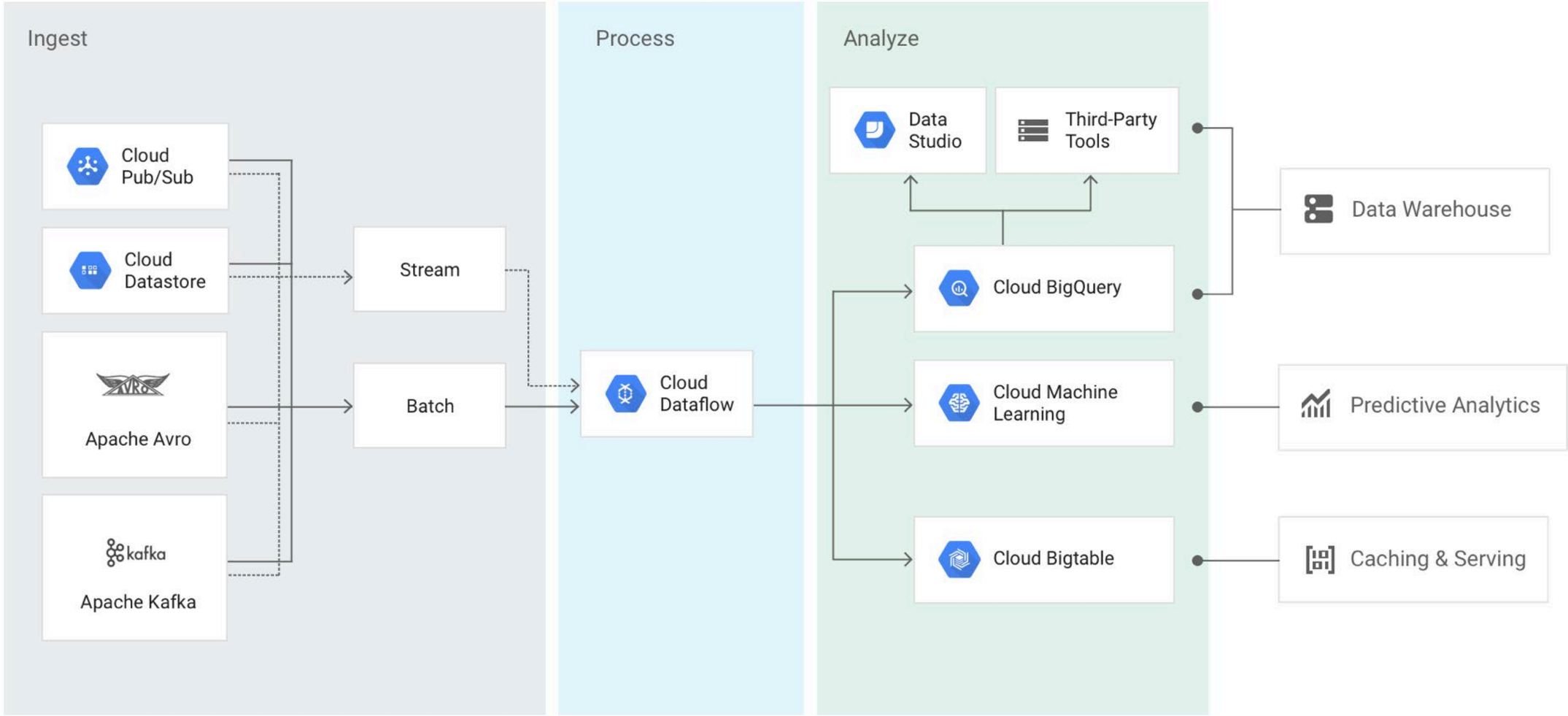
Cat vs. Dog Image Classification

20 mins

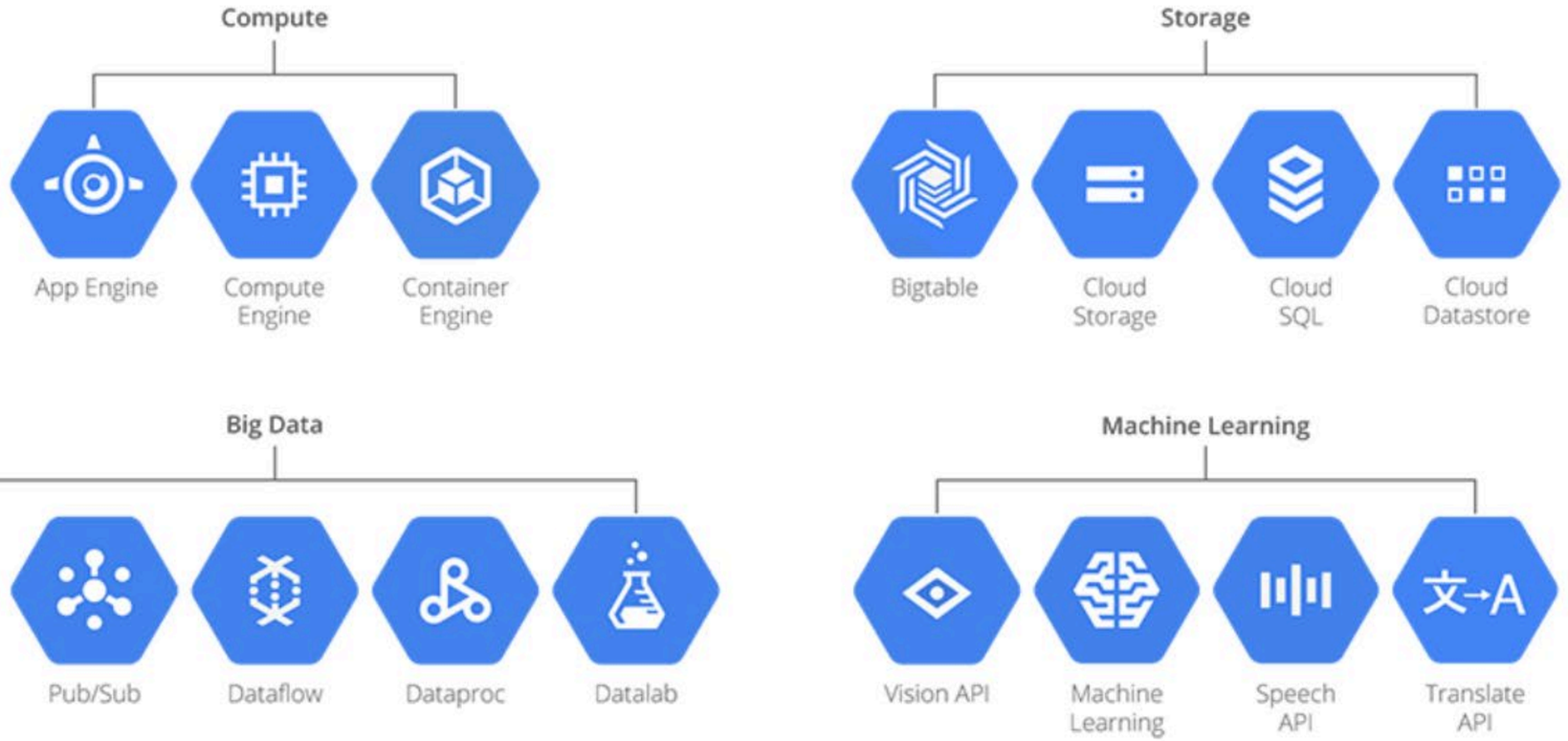
https://developers.google.com/machine-learning/practica/image-classification/

Cloud ML

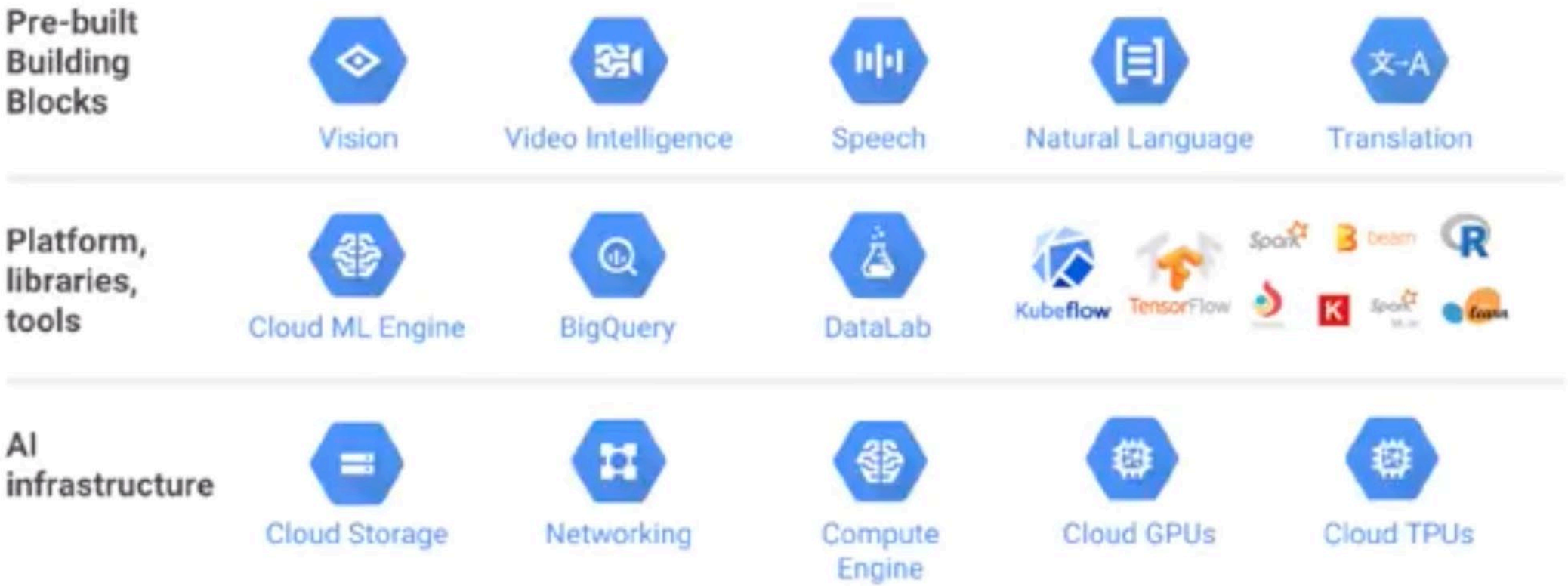
Google Cloud Platform



Google Cloud Platform

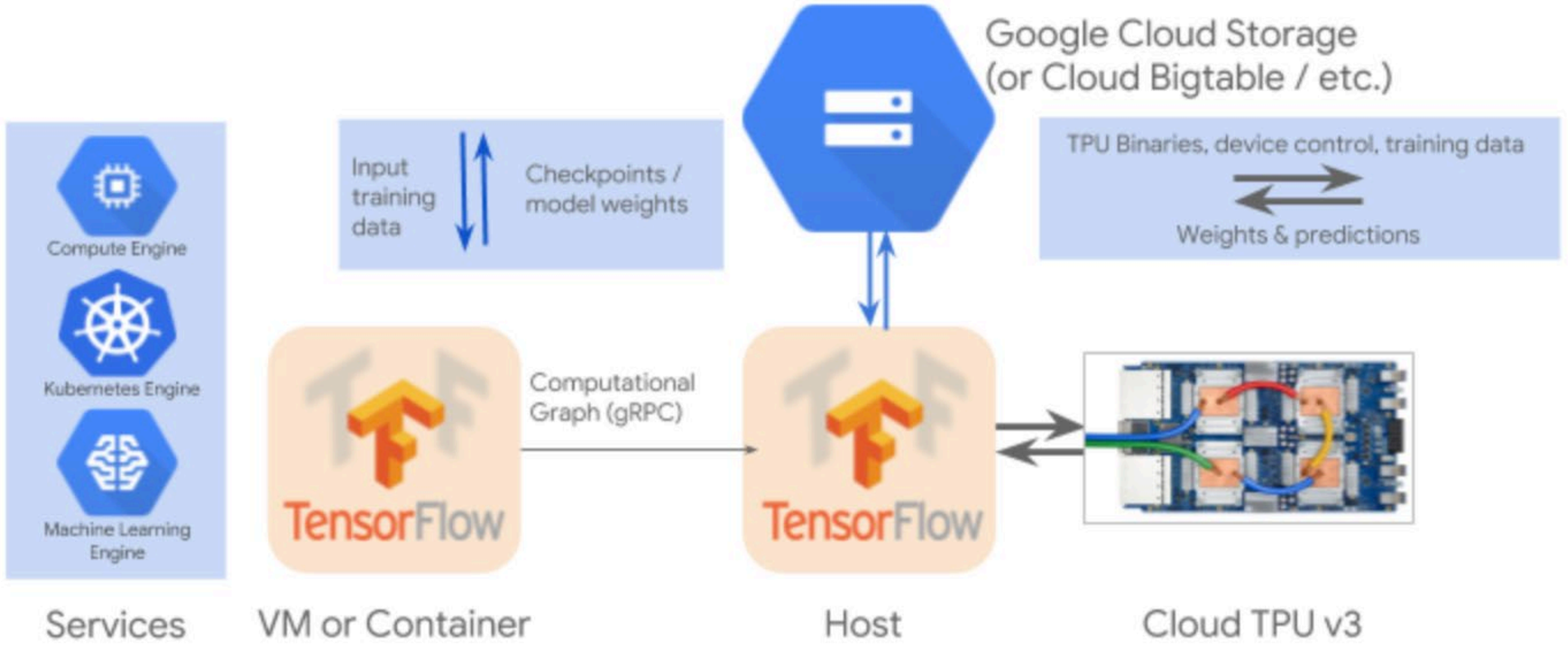


Google Cloud AI Platform



https://cloud.google.com/products/ai/

Cloud TPU



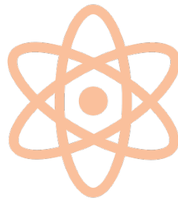
Use Cases

- Finance
- Healthcare
- Self-driving cars

Finance

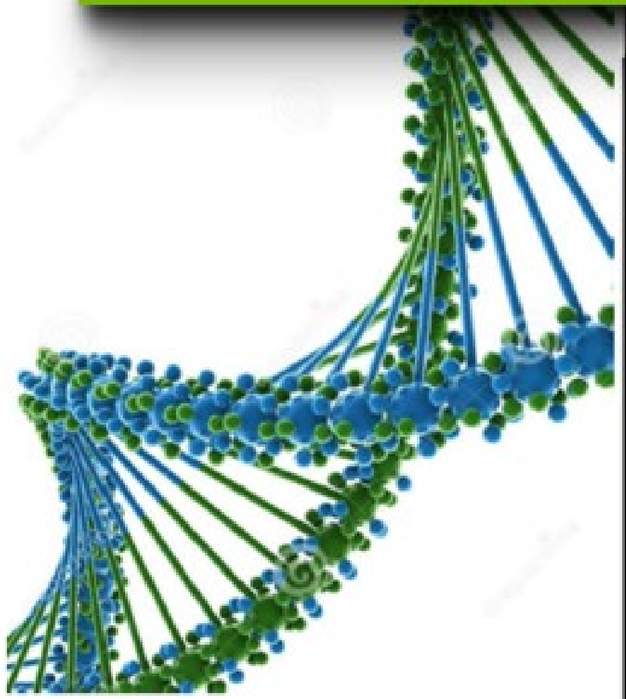
Finance – Use Cases

- **Cybersecurity** - Anomaly (Fraud) Detection. Can detect and predict irregular activity
- **Reporting** - systems that can instantaneously analyze data, reason about what that analysis means and then generate natural language reports that inform many audiences. Useful in compliance and regulation
- **Risk Management** - insurance, trading, minimizing losses
- **Advising** - system has a "conceptual awareness" of client goals and needs, robo-advisors
- **Trading** - Deep Learning is being used in algorithmic, high frequency trading. Also sentiment analysis
- **Portfolio Optimization** - Ideal for creating and optimizing investment portfolio

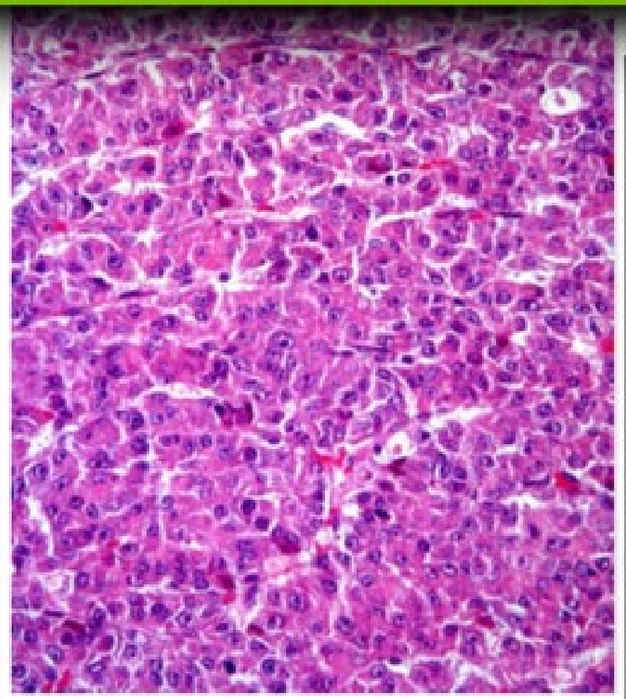


Healthcare

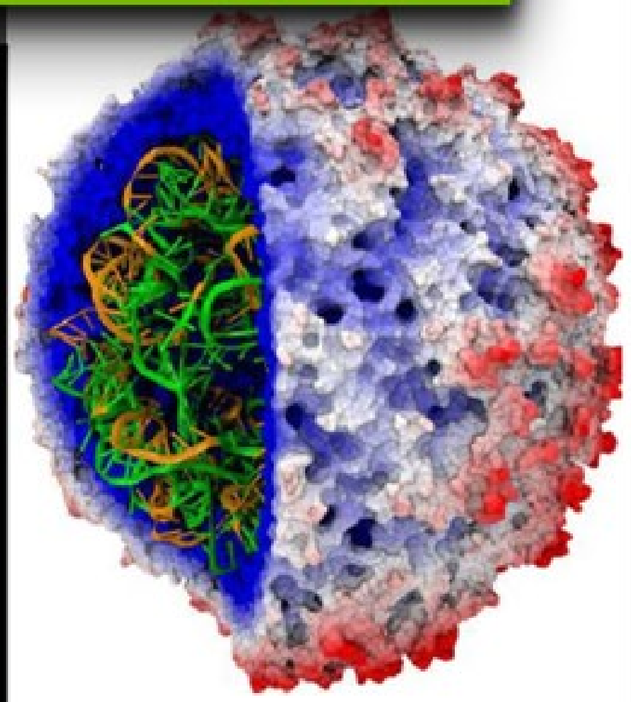
AI For Cancer Moonshot
Big Data & Precision Oncology



Discover
Cancer's Genetic Signatures



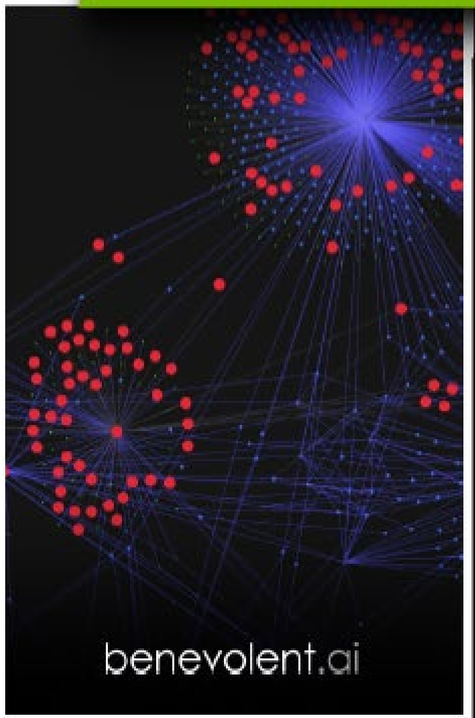
Automate Pathology
to Find Cancer Biomarkers



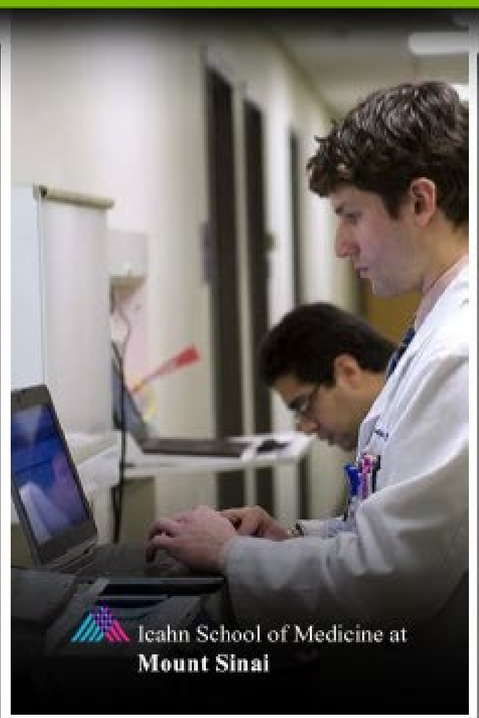
Predict Cancerous
Protein Formations

June 2016 NCI Genomic Data Commons = 3PB Genomics Data

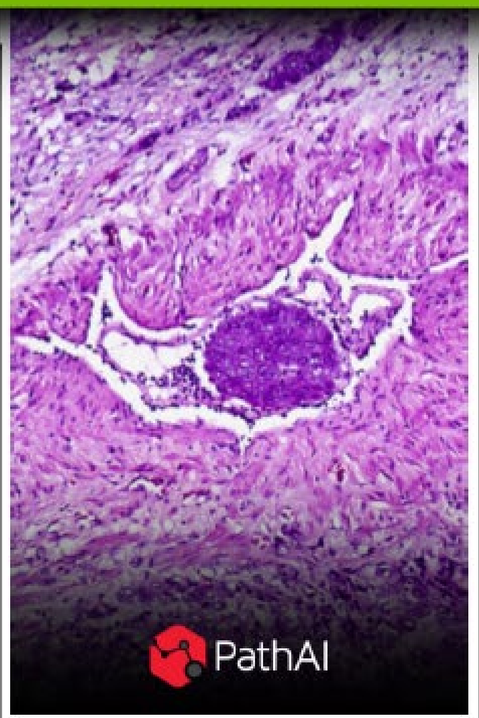
Where else is GPU Deep Learning Impacting Healthcare?



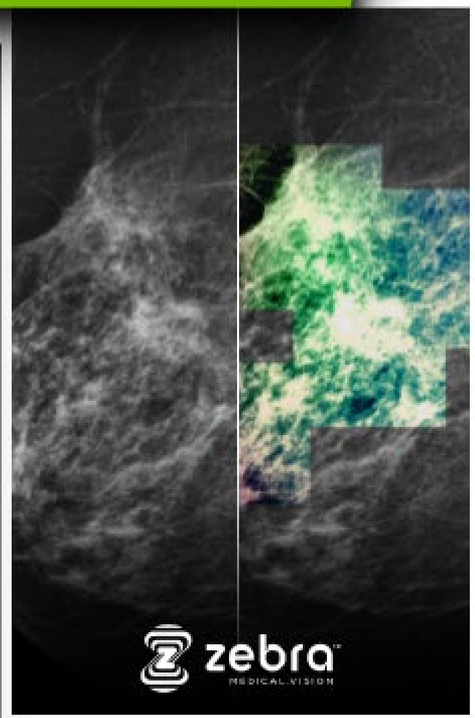
Accelerating Targeted Drug Development



Predicting Disease from Medical Records

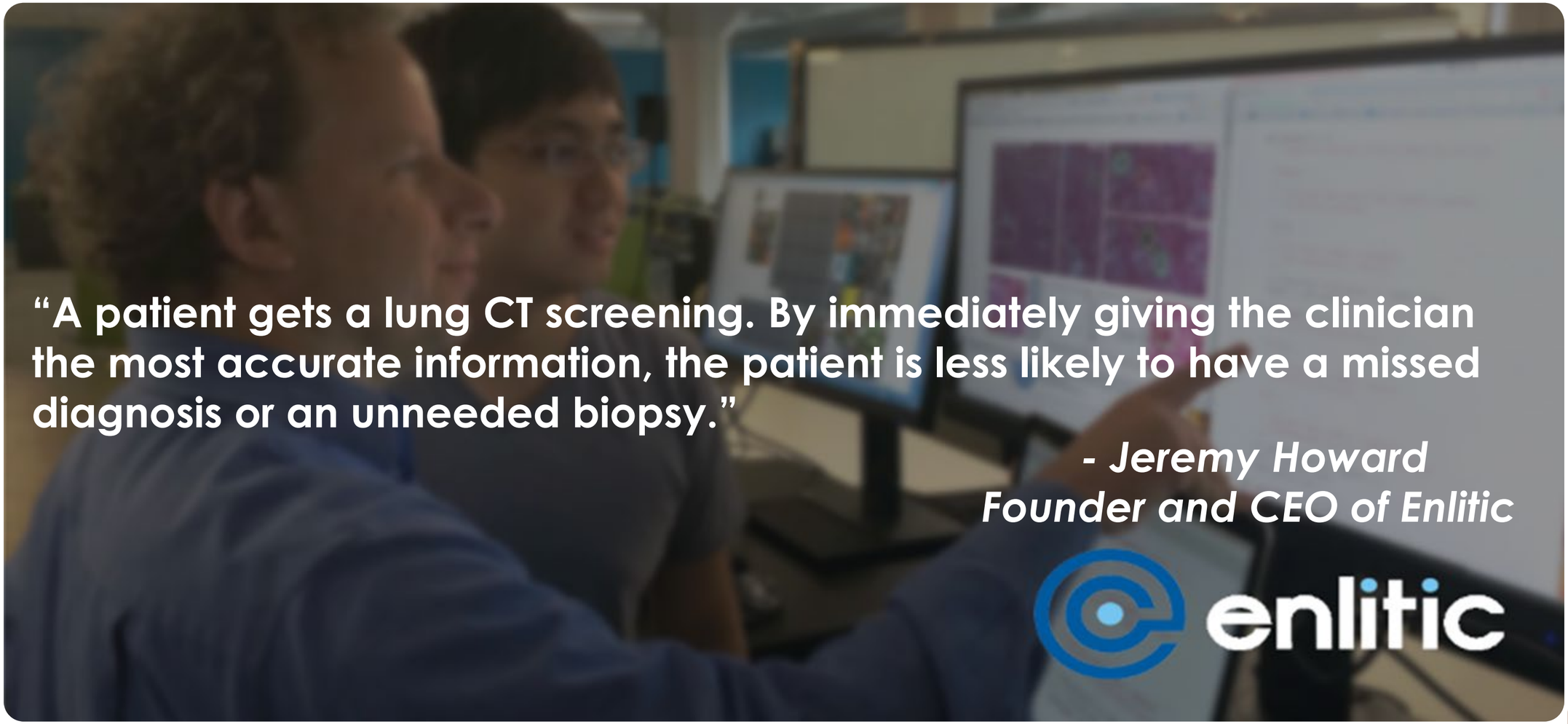


Reducing Cancer Diagnosis Error Rate by 85%



Teaching Computers to Recognize Cancer

Enlitic applies state of the art Deep Learning technology to medicine

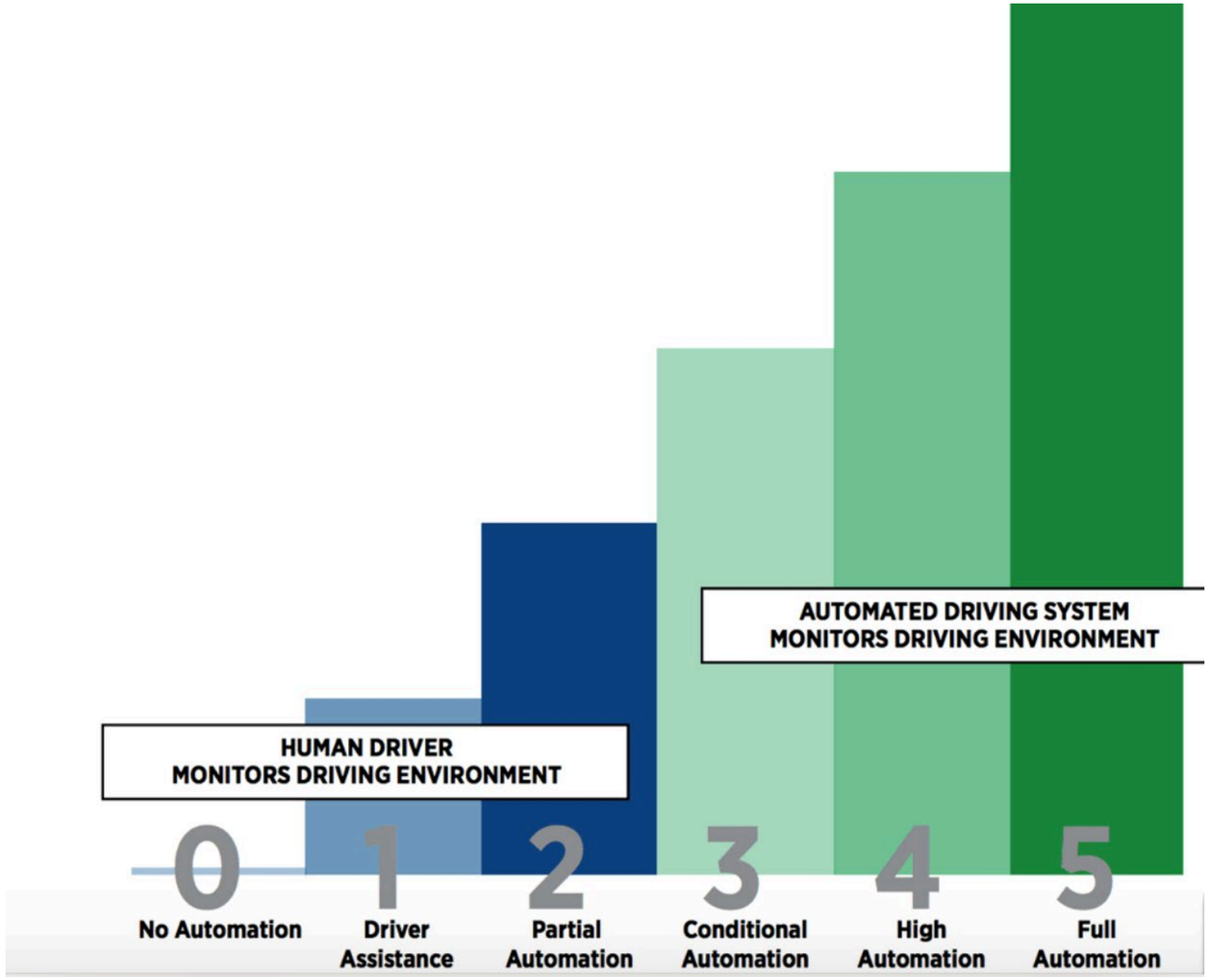


Insilico Medicine applies Deep Learning algorithms to drug discovery

Pathway Genomics developed a blood test kit with IBM Watson to test early cancer detection

Self-driving Cars

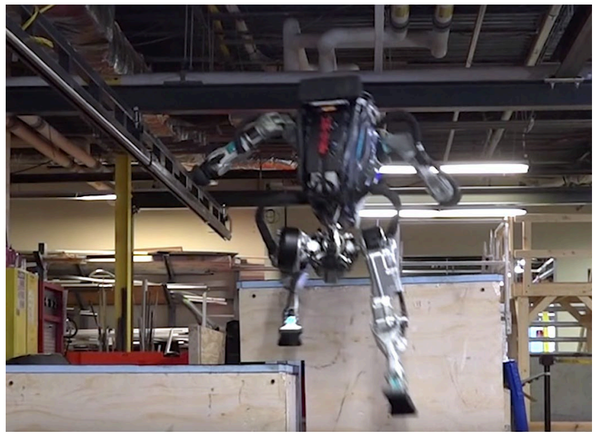
Six Levels of Autonomous Driving



Self-driving Cars

- All major car manufacturers (Ford, Mercedes, Toyota, …)
- All processor companies (Intel, Nvidia, AMD, …)
- New players (Tesla, Aurora, Waymo, Uber, Oxbotica, Drive.ai)
- Online courses – Udemy, MIT

Robotics



Robotics

Robotics

- Use Cases – Industrial, Domestic, Transportation, Warehouse, …
- Companies – Kuka, Dyson, iRobot, ABB, Fanuc, Fetch, Softbank, …
- Research – Berkeley, MIT, CMU, Imperial, …
- Technology – Reinforcement learning, control systems

THANK YOU
WE HOPE YOU ENJOYED THIS COURSE