Data Science/Machine Learning/Deep Learning Workshop

Nueva Ecija University of Science and Technology (NEUST)

Rodolfo C. Raga Jr., PhDCS

December 21-23, 2021

Day 1 : Basic Concepts, Intuitions, and Foundations

Topic 1: Introduction to Data Science, Machine Learning, and Deep Learning
- What is Data Science
- What is Machine Learning
- What is Deep Learning

Topic 2: Intro to Python, Jupyter Notebook, Google Colab, Scikit-learn, Tensorflow, and Keras
- Why use Python
- Jupyter Notebook vs. Google Colab
- Understanding sklearn, Tensorflow and Keras

Topic 3: Fundamentals of Data Analysis using sklearn
- Correlation and Distribution Analysis
- Data Preprocessing (Feature Selection, Data Normalization, Data Splitting)
- Building and Training of ML prediction models using sklearn

Topic 4: Fundamentals of Neural Networks
- Definition and NN Architecture
- Perceptron and Multi-layer Perceptron Architecture
- Building NN prediction models using sklearn

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Outline

◈ Data Science, Why is it a thing?

◈ Artificial Intelligence, How smart is it?

◈ Machine Learning, What does it do?

◈ Deep Learning, Where can it go?

◈ Applications Demo :

➡ Predicting Student Performance using Various ML algorithms

➡ Using Python for Image Classification

➡ Customized image classification using CNN

Companies collect huge Volumes of data

◈ Google processes 20 PB a day

◈ Facebook has 60 TB of daily logs

◈ eBay has 6.5 PB of user data + 50 TB/day

◈ Cost of 1 TB of disk: P3,000.00

◈ Time to read 1 TB disk: 3 hrs
  (100 MB/s)
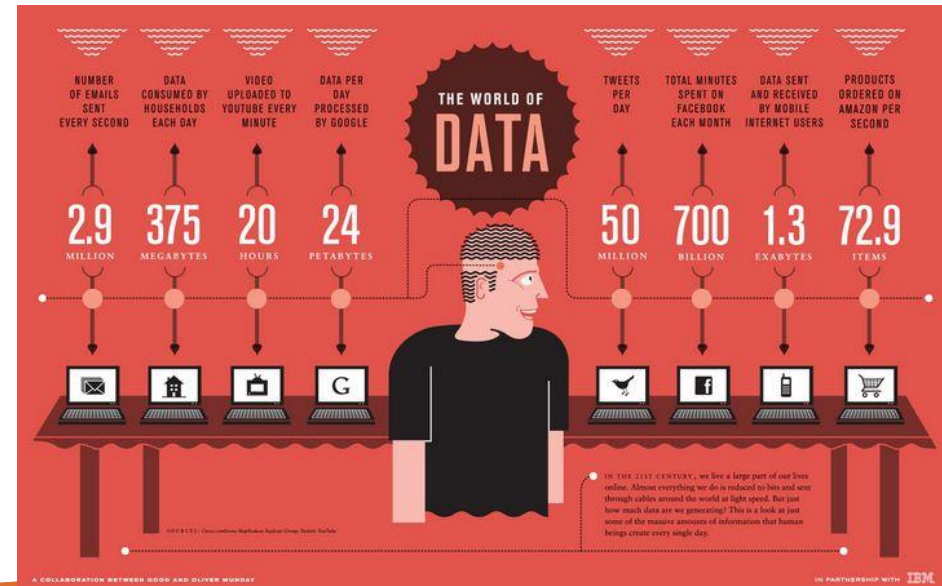
Flood of data from Users and Devices

By 2020

Avg. internet user	1.5 GB of traffic / Day
Autonomous vehicles	4 TB of data / Day*
Connected airplane	5 TB of data / Day
Smart Factory	1 PB of Data / Day
Cloud video Providers	750 pB of video / Day

The Coming Flood of Data

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

The emergence of Big Data

◈ Big Data is any data that is expensive to manage and hard to extract value from

- ➡ **Volume**
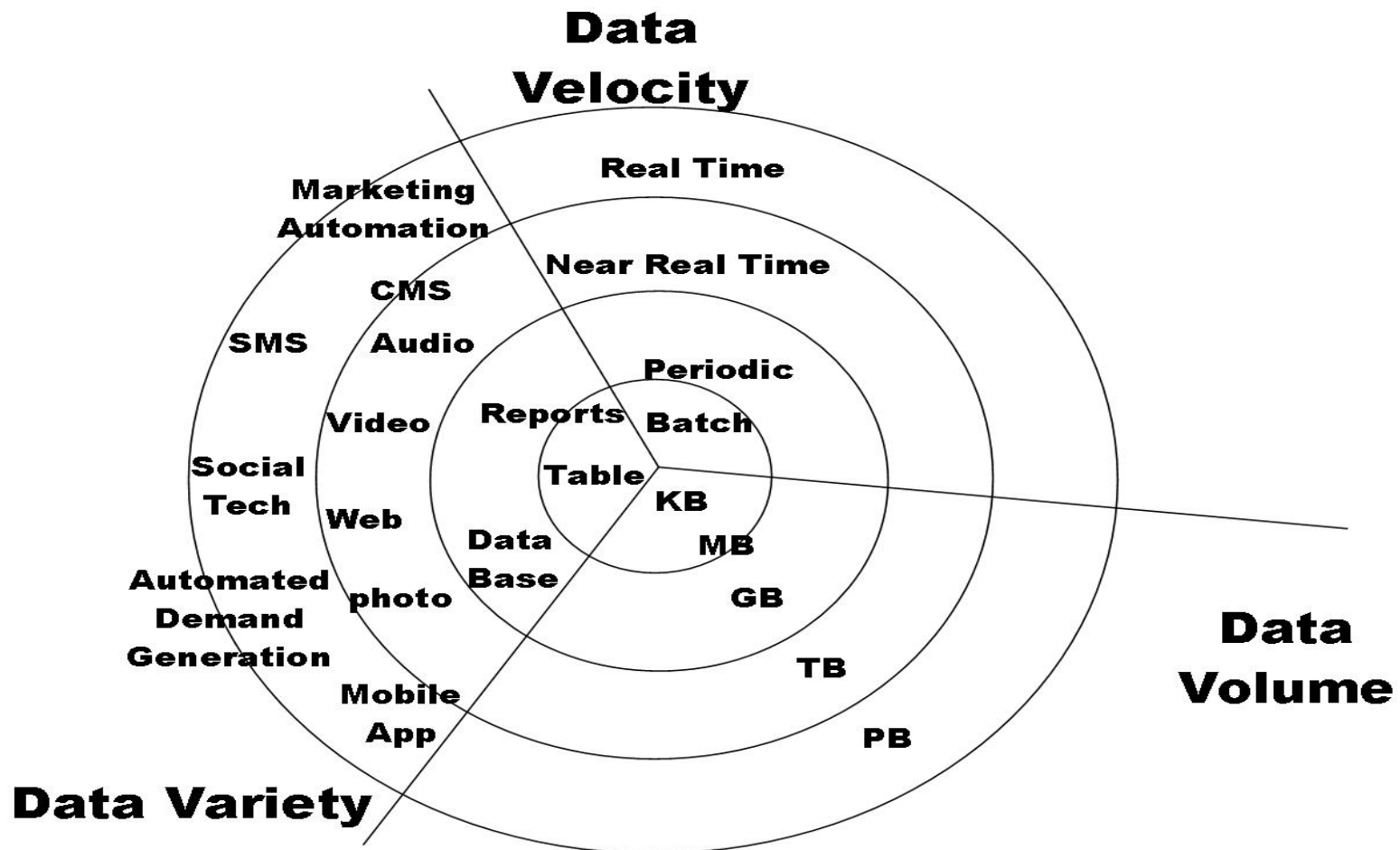 - ◆ The size of the data
- ➡ **Velocity**
  - ◆ The latency of data processing relative to the growing demand for interactivity
- ➡ **Variety and Complexity**
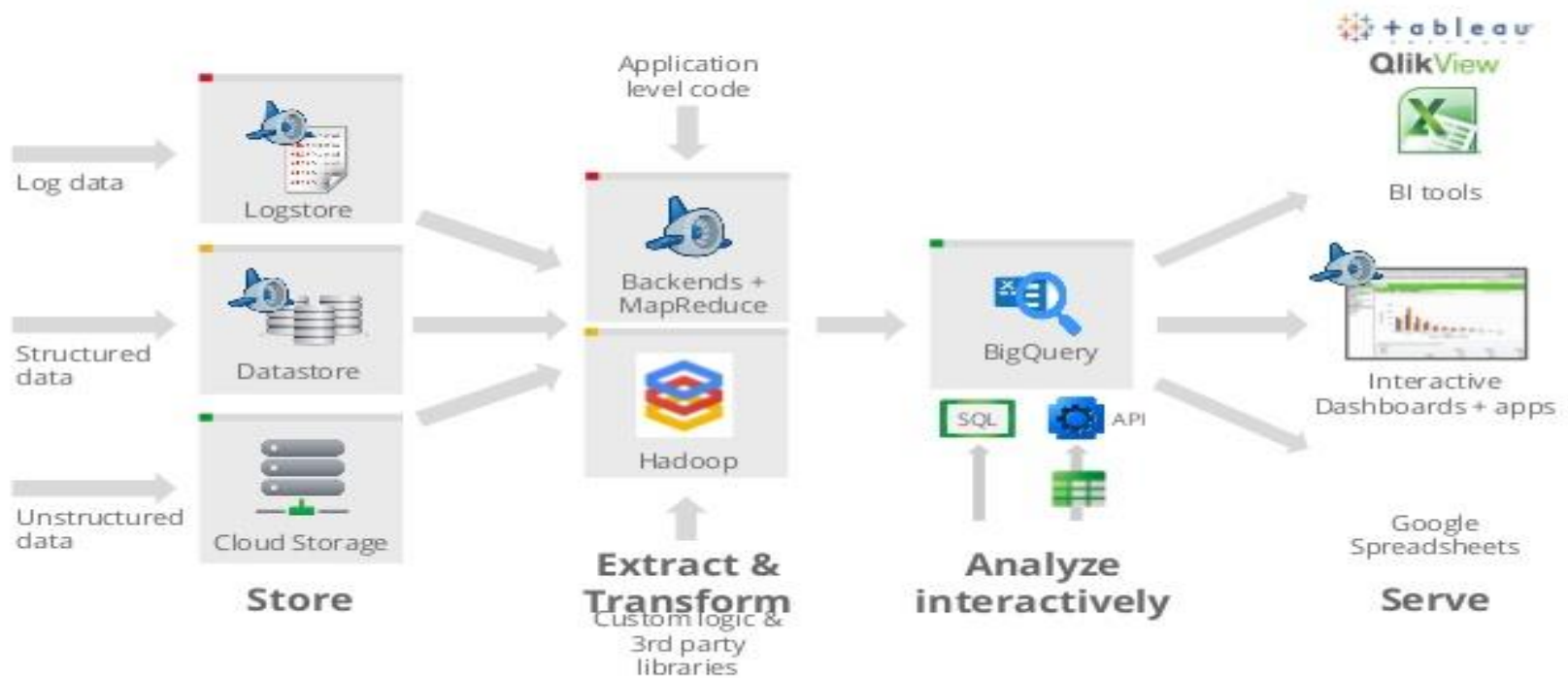 - ◆ the diversity of sources, formats, quality, structures.

Realms of Big Data

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Big Data Need Processing to be Useful

Big Data Processing Pipeline

This is where Data Science Comes in…

- Data Science is an area that manages, manipulates, extracts, and interprets knowledge from tremendous amount of data

- Data science (DS) is a multidisciplinary field of study with goal to address the challenges in big data

- Data science principles apply to all data – big and small

Why Data Science is a concern?

$16.9 Billion Industry

DEMAND FOR BUSINESS INTELLIGENCE AND ANALYTICS HITS TIPPING POINT

Home / Data Center / Demand for Business Intelligence and Analytics Hits Tipping Point

Gartner, known worldwide for its research and projections, predicted several years ago that data center managers would be swamped with a flood of demands for resources related to business intelligence and analytics.

It appears that time has arrived, with Gartner recently announcing that the shift from the old way of things to the more modern business intelligence and analytics platform has reached "a tipping point."

The numbers back up that statement. The research firm reported that global revenue in the business intelligence and analytics market is forecast to reach $16.9 billion in 2016 — up 5.2 percent from 2015.

"Organizations must transition to easy-to-use, fast and agile modern BI platforms to create business value from deeper insights into diverse data sources," said Ian Bertram, managing vice president at Gartner.

The way Gartner puts it, that development represents an ongoing transition from analytics gathered from IT-led system-of-record reporting to business-led self-service analytics. It's all designed to make the process of accessing data easier, faster and more agile under the modern BI and analytics platform. As a

LIFELINE ONE SHEET

Want a quick look at what we do and who we are? Download our Lifeline Data Centers One Sheet.

DOWNLOAD NOW

COMPARTMENTALIZATION WHITE PAPER

We've created a comprehensive guide on data center power compartmentalization and why it's

Demand for Data Scientist

"By 2018, there will be a shortage of **1.5 MILLION** analysts/managers who can make data-driven decisions."
- McKinsey Global Institute's report on big data

"The analytics market in India could more than double from the current $1 billion to $2.3 billion by the end of 2017-18" **$2.3 BILLION**
Industry report by NASSCOM – in partnership with BlueOcean Market Intelligence

Analytics firms in India will soon face a shortage of **2 LACS** data scientists
- Hindu

The worldwide business analytics market will grow from $37.7B in 2013 to **$59.2 BILLION** in 2018, attaining a 9.4% CAGR in the forecast period.
- IDC (International Data Corporation)

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Why Data Science is a Critical Skill to learn

CIO   ASEAN ▼   DIGITAL MAGAZINE   EVENTS & AWARDS PROGRAMS   VIDEO   NEWSLETTERS   RE

The 13 most in-demand tech jobs for 2019

Job	25th percentile	50th percentile	75th percentile	95th percentile
Business intelligence analyst	$85,750	$106,000	$132,000	$178,000
Cloud architect	$75,000	$94,500	$118,000	$159,500
Cloud systems engineer	$86,250	$103,000	$123,250	$145,750
Data scientist	$102,750	$121,500	$147,500	$175,000
Database developer	$98,250	$118,000	$141,000	$167,750
Developer (web, software, mobile)	$83,500 (web); $98,250 (software); $65,600 (mobile)	$100,250 (web); $117,500 (software); $79,000 (mobile)	$119,750 (web); $140,750 (software); $93,500 (mobile)	$142,000 (web); $166,500 (software); $105,000 (mobile)
DevOps engineer	$90,250	$110,500	$134,750	$178,250
Full-stack developers	$65,000	$79,250	$96,000	$130,500
Help desk and desktop support specialists	$49,000 (tier 1); $38,250 (tier 2); $32,250 (tier 3)	$58,500 (tier 1); $45,740 (tier 2); $54,750 (tier 3)	$70,000 (tier 1); $54,750 (tier 2); $46,000 (tier 3)	$83,750 (tier 1); $64,500 (tier 2); $55,000 (tier 3)

Nature of Data Science

Data Science Is Multidisciplinary

By Brendan Tierney, 2012

DS vs AI vs ML vs DL

DATA SCIENCE
Concerned with extracting insights from data in its various forms. Although it has no direct relation to AI, ML or DL, it can be useful to each of the three.

ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Artificial Intelligence

Artificial Intelligence is centered on the design and implementation of :

- Programs that can make smart decisions
  - at least for a narrowly defined problem

- Machines that have perception (of their environment) and can act on it

- Systems that can mimic human cognition

- Agents that are:
  - Rational: act to maximize a quantitative and objective performance measure.
  - Autonomous: able to operate without guidance.
  - Adaptable: able to improve with experience.

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

AI – How smart can it get?
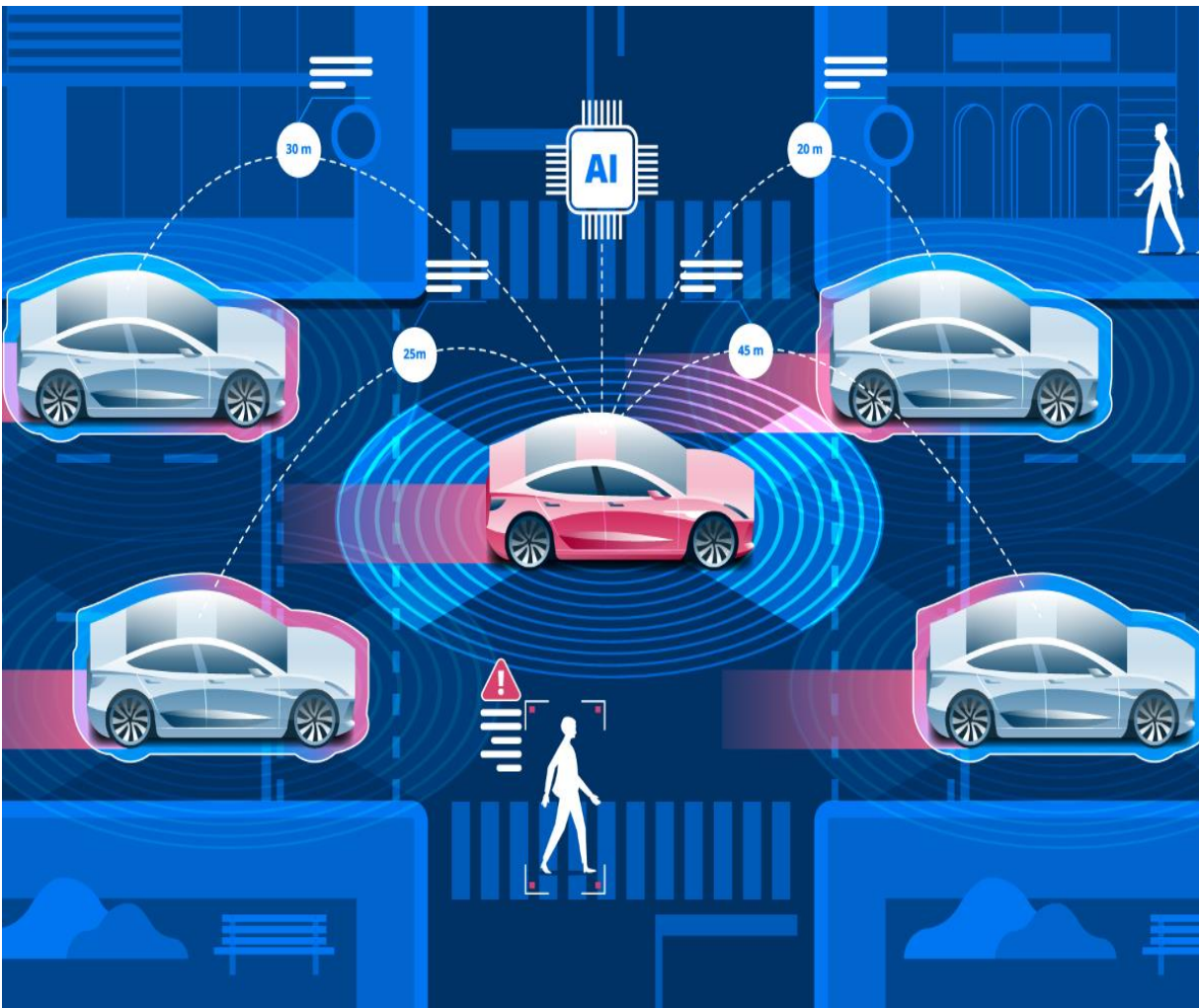
A **self-driving car**, also known as an autonomous car or driverless car is a vehicle that is capable of sensing its environment and moving safely with little or no human input.

Self-driving cars combine a variety of sensors to perceive their surroundings, such as radar, lidar, sonar, GPS, odometry and inertial measurement units.

Advanced control systems interpret sensory information to identify appropriate navigation paths, as well as obstacles and relevant signage.

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

AI – How smart can it get?

WAYMO LAUNCHES FIRST US COMMERCIAL SELF-DRIVING TAXI SERVICE

Service limited to city of Phoenix - for now

Andrew Buncombe Seattle | @AndrewBuncombe |
Wednesday 5 December 2018 22:50 |

Almost ten years after Google secretly started work on technology that would allow a vehicle to operate without a human driver, the company has launched the nation's first commercial self-driving robo-taxi.

Waymo, a subsidiary of Google, introduced a small fleet of ride-hailing vehicles in Phoenix, Arizona, asking people to pay, just as they would to travel by Uber or Lyft. For now, the project will also feature a human driver behind the wheel, just in case the robotic vehicle malfunctions.

"Over time, we hope to make Waymo One available to even more members of the public Self-driving technology is new to many, so we're proceeding carefully," Waymo's CEO John Krafcik, wrote in a blog post about Wednesday's run-out.

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

# AI – How smart can it get?

- Nevada made it legal for autonomous cars to drive on roads in June 2011

- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

Penn's Autonomous Car → (Ben Franklin Racing Team)
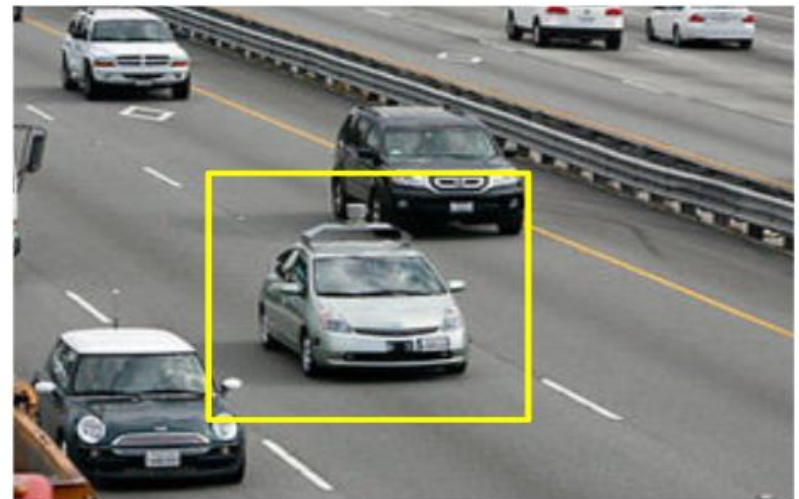
AI – How smart can it get?

Which States Allow Self Driving Cars?

Legend:
- Legal (green)
- No Legislation (dark blue)
- Legal w/ Safety Driver (orange)
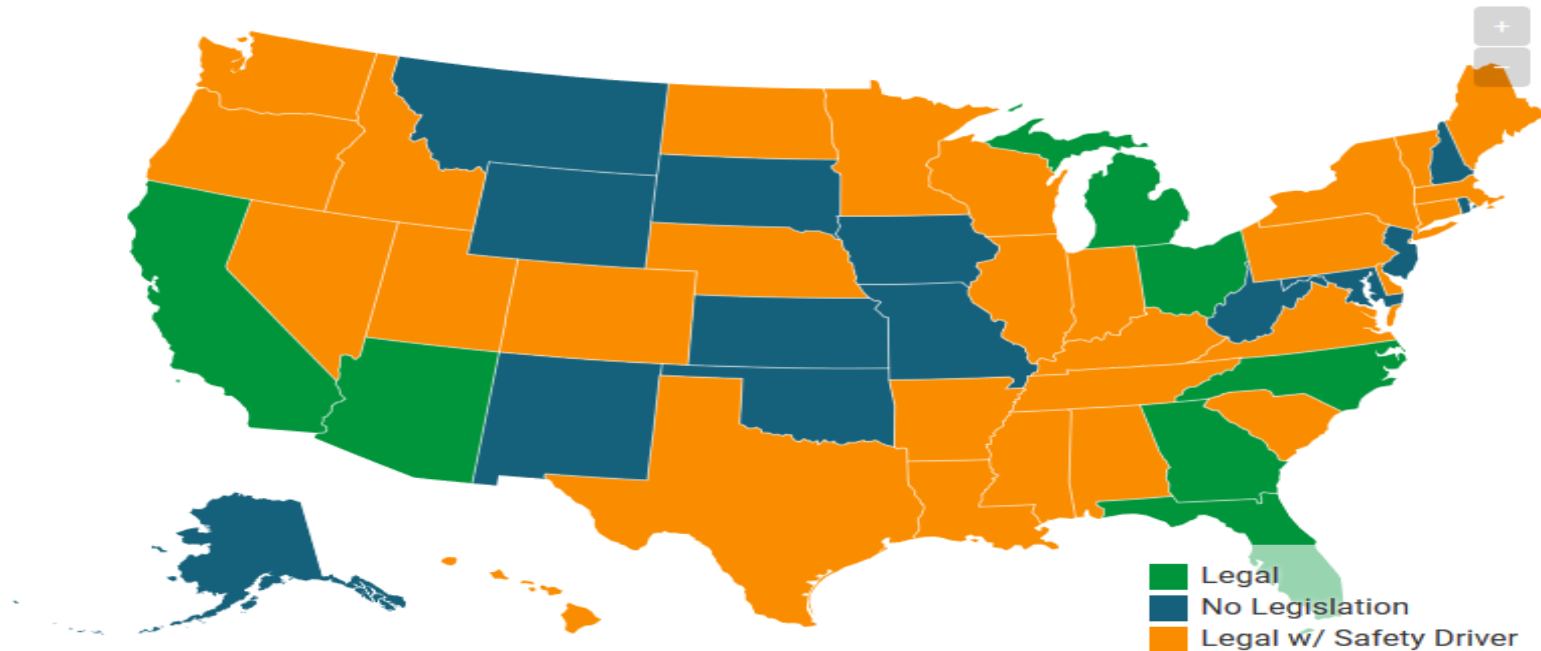
Map: Lifewire.com • Source: State Legislative Offices

States that allow fully autonomous self-driving car testing on public roads:

Arizona, California, Florida, Georgia, Michigan, North Carolina, Ohio

States that allow self-driving car testing with a safety driver:

Alabama, Arkansas, Colorado, Connecticut, Delaware, Hawaii, Idaho, Illinois, Indiana, Kentucky, Louisiana, Maine, Massachusetts, Minnesota, Mississippi, New York, North Dakota, Nebraska, Nevada, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Utah, Vermont, Virginia, Wisconsin, Washington, Washington D.C.

AI – How smart can it get?

Death of Elaine Herzberg

The **death of Elaine Herzberg** (August 2, 1968 – March 18, 2018) was the first recorded case of a pedestrian fatality involving a self-driving (autonomous) car, following a collision that occurred late in the evening of March 18, 2018. Herzberg was pushing a bicycle across a four-lane road in Tempe, Arizona, United States, when she was struck by an Uber test vehicle, which was operating in self-drive mode with a human safety backup driver sitting in collision, Herzberg was taken to the hospital where she died of her injuries.[2][3][4]

As a result of the fatal incident, Uber immediately suspended testing of self-driving vehicles in Arizona,[5] where such testing had 2016.[6] Uber also decided not to renew its permit for testing self-driving vehicles in California when it expired at the end of March

Herzberg was specifically the first *pedestrian* death involving a self-driving car; a previous fatality, in which the occurred almost two years prior.[8] A *Washington Post* reporter compared Herzberg's fate with that of Bridget first pedestrian to be killed by an automobile.[9] This incident has magnified the importance of collision avoida
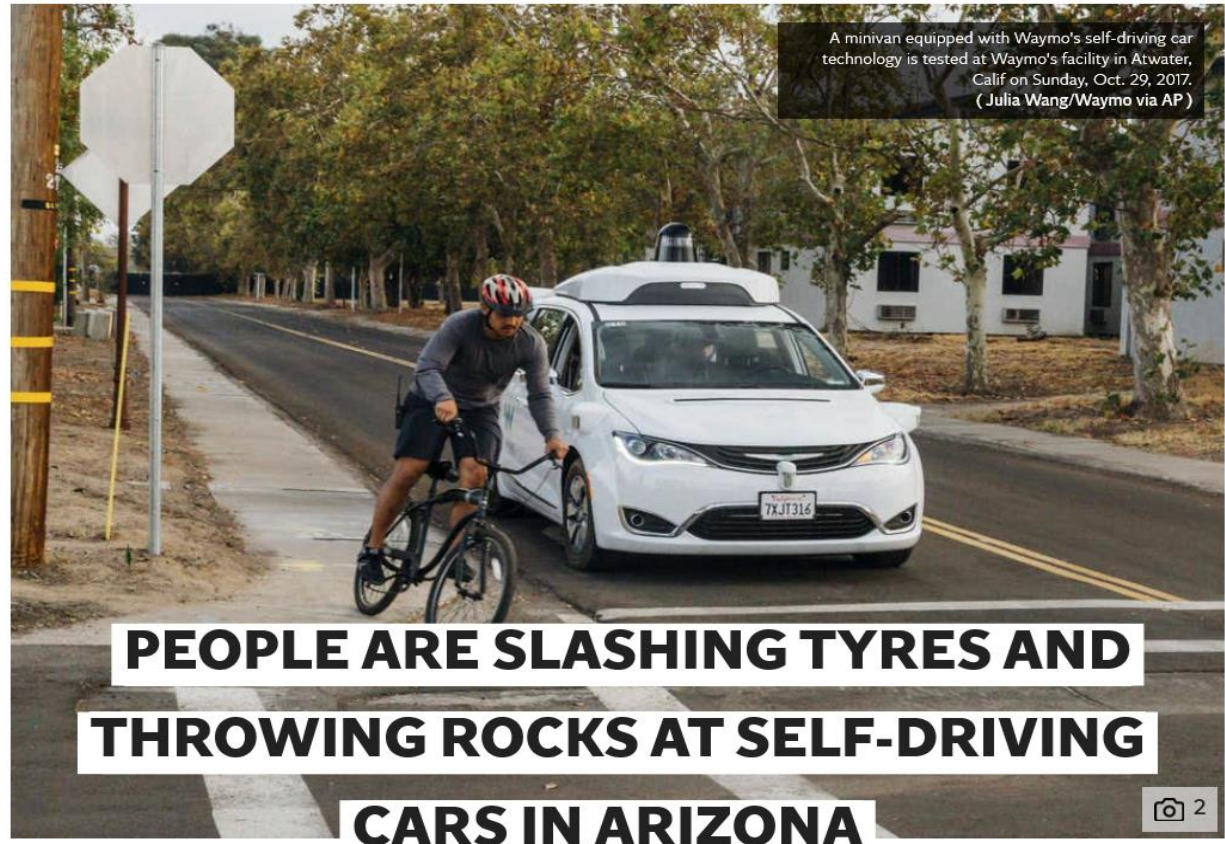

1
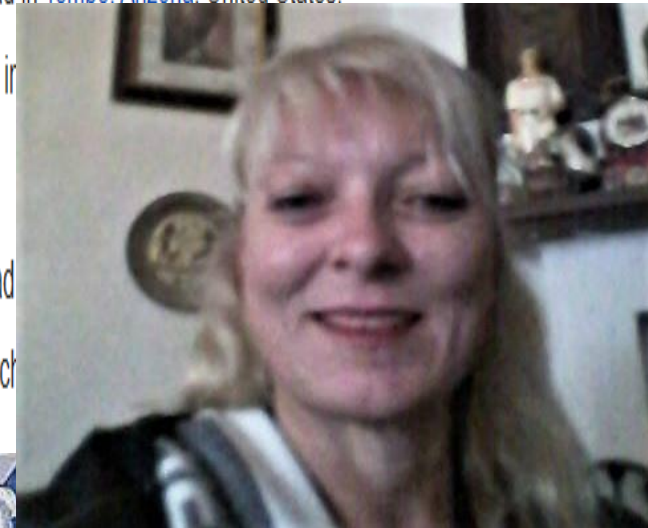
AI – How smart can it get?

Vigilante citizens in a town in Arizona have slashed tyres, thrown rocks and even pointed guns at self-driving vehicles being tested in their neighbourhood, an investigation has revealed.

Police in Chandler recorded 21 incidents over the past two years in which the autonomous vehicles and their test drivers were targeted by local residents.

One incident on 24 October saw a man emerge from a park and slash the tyres of a Waymo vehicle stopped at an intersection. Earlier this year a Waymo test driver reported a man in shorts aiming a gun at his car when it passed the man's driveway.

CULTURE · VIDEO · INDY/LIFE · HAPPY LIST · INDYBEST · LONG READS · INDY100 · VOUCHERS · PREMIUM

A minivan equipped with Waymo's self-driving car technology is tested at Waymo's facility in Atwater, Calif on Sunday, Oct. 29, 2017.
(Julia Wang/Waymo via AP)

PEOPLE ARE SLASHING TYRES AND THROWING ROCKS AT SELF-DRIVING CARS IN ARIZONA

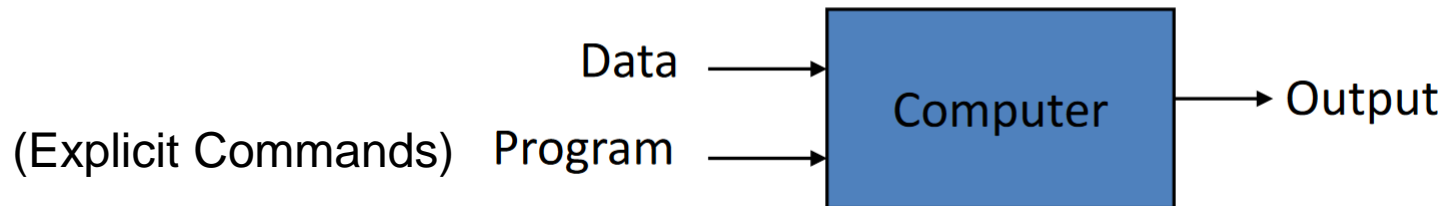Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Machine Learning – What does it do?

- **Machine Learning** is a branch of artificial intelligence (AI) whose goal is to develop systems with abilities to automatically learn and improve from experience (data) without being explicitly programmed.

- **Machine Learning** is achieved by enabling computer programs to process data and use it to find patterns that they can learn from for themselves.
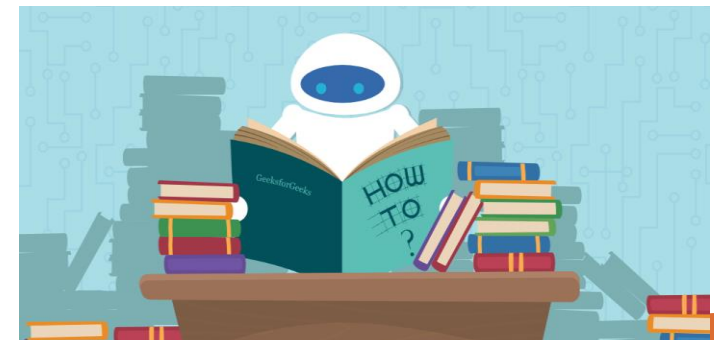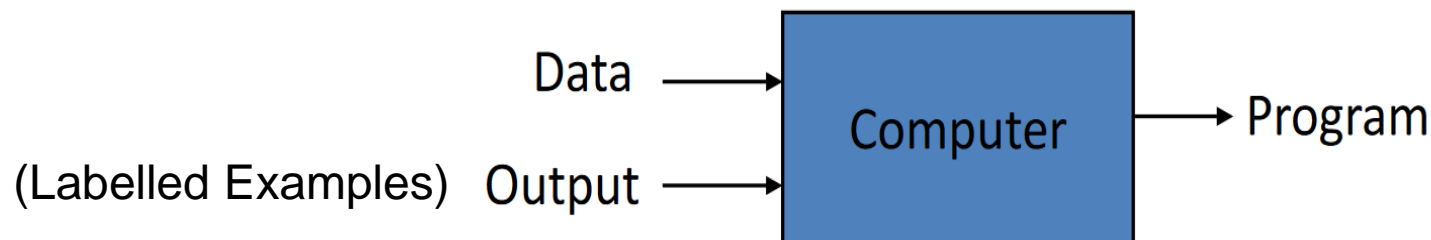
Traditional Programming vs Machine Learning

Traditional Programming

Data $\longrightarrow$ Computer $\longrightarrow$ Output

(Explicit Commands) Program $\longrightarrow$

Machine Learning

Data $\longrightarrow$ Computer $\longrightarrow$ Program

(Labelled Examples) Output $\longrightarrow$

Machine learning tasks

◆ Supervised learning
- **Input**: training data + desired outputs (labels)
- regression: predict numerical values
- classification: predict categorical values, i.e., labels

◆ Unsupervised learning
- **Input**: training data (without desired outputs)
- clustering: group data according to "distance"
- association: find frequent co-occurrences
- link prediction: discover relationships in data
- data reduction: project features to fewer features

Supervised Machine learning

- Cortana and other digital assistants such as SIRI and Google Assistant, are essentially speech automated systems in mobile phones or laptops. This software responds with an appropriate answer every time you ask it a question.

- The way it works is by collecting a bunch of questions along with their answers. Using these, it trains a model to create a question-answer software.

- They train themselves to work with your inputs and then deliver amazing results according to their training.
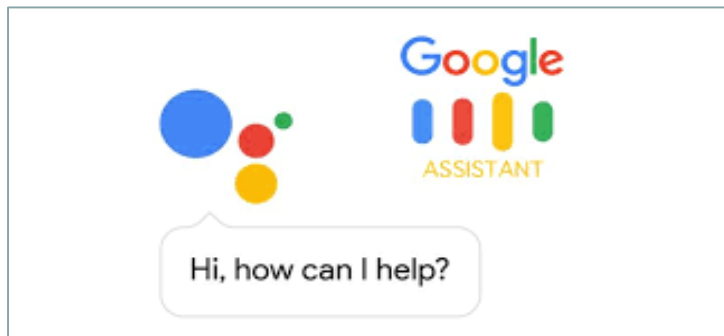
Unsupervised Machine learning

Customer Segmentation Clustering
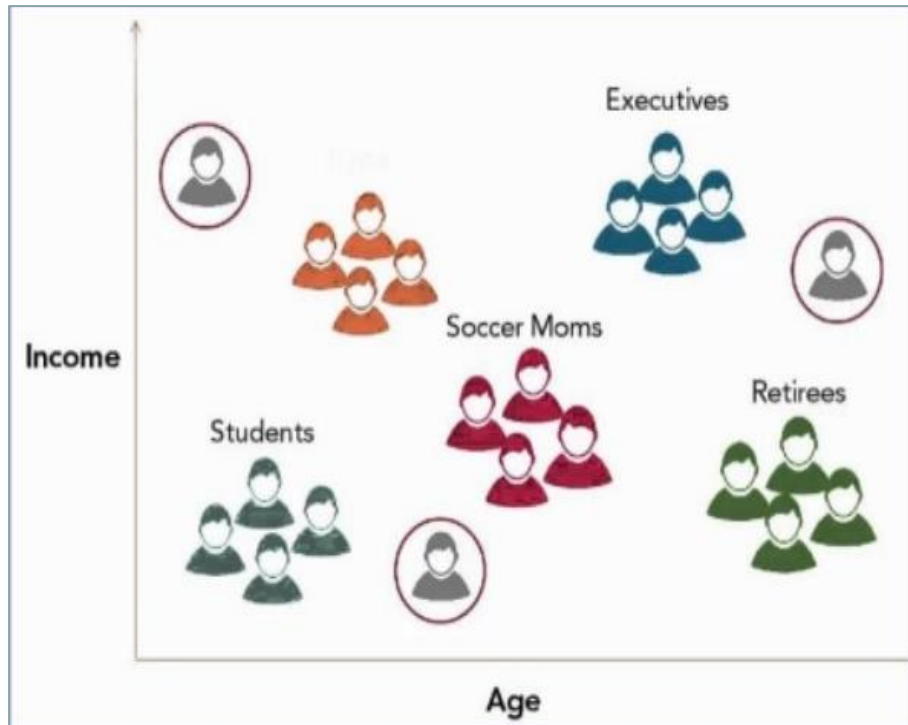
What products are bought together (basket analysis)
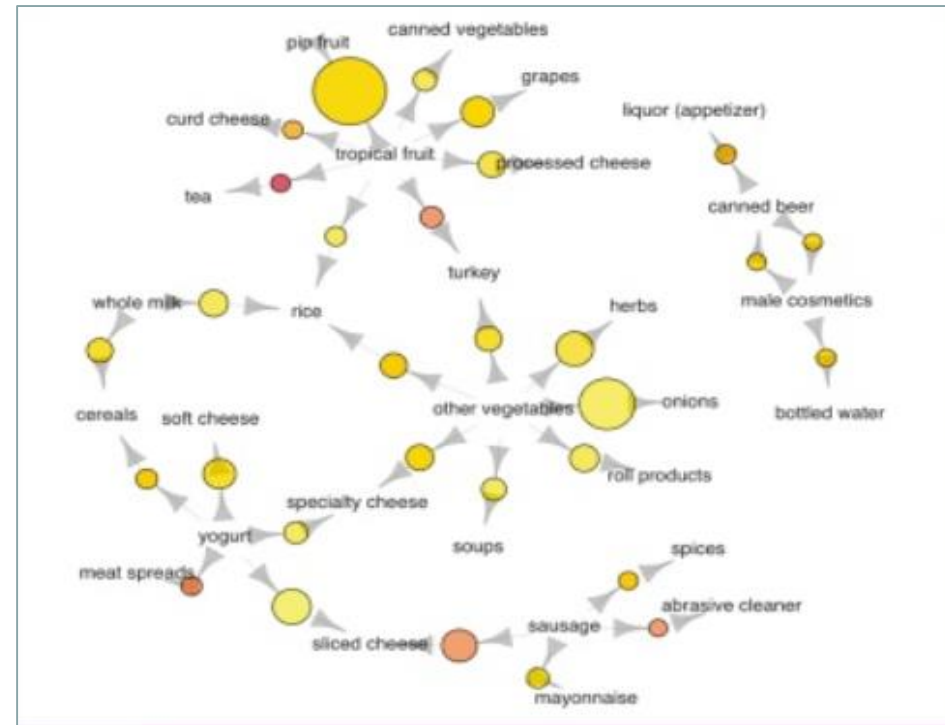
Other Engineering Applications of ML

Manufacturing
- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Propensity to buy
- Demand forecasting
- Process optimization
- Telematics

Retail
- Predictive inventory planning
- Recommendation engines
- Upsell and cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value

Healthcare and Life Sciences
- Alerts and diagnostics from real-time patient data
- Disease identification and risk stratification
- Patient triage optimization
- Proactive health management
- Healthcare provider sentiment analysis

Travel and Hospitality
- Aircraft scheduling
- Dynamic pricing
- Social media – consumer feedback and interaction analysis
- Customer complaint resolution
- Traffic patterns and congestion management

Financial Services
- Risk analytics and regulation
- Customer Segmentation
- Cross-selling and up-selling
- Sales and marketing campaign management
- Credit worthiness evaluation

Energy, Feedstock, and Utilities
- Power usage analytics
- Seismic data processing
- Carbon emissions and trading
- Customer-specific pricing
- Smart grid management
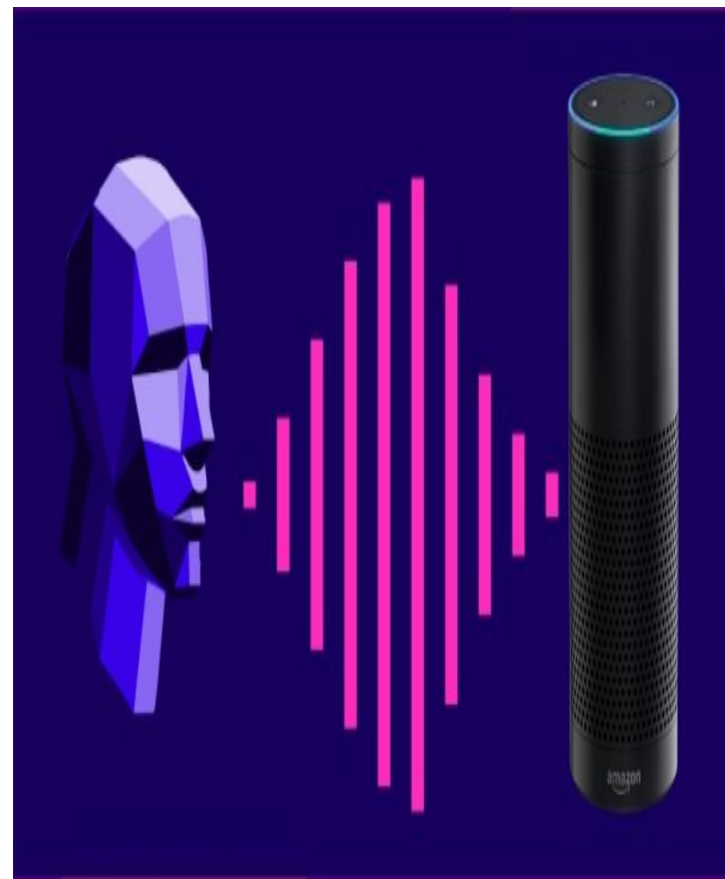- Energy demand and supply optimization

Source: TATA Consultancy Services

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Applications with varied Data Modality

- Image Recognition

- Speech Recognition

- Natural Language Processing

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Deep Learning: A Revolution in ML

- **Deep learning** is only a subset of machine learning. It is one of the most popular forms of machine learning.
- They use **Artificial Neural Networks (ANNs)**. Artificial neural networks are similar to biological neural networks in humans. It consists of layers of "neurons" connected to each other.
- **Neurons** are interconnected and can communicate with each other. It combines many inputs and produces an output using a simple math function.
- Final outputs depend on the **weights of different neurons.**
- # of layers = depth of the network. Hence the term **Deep Learning**.
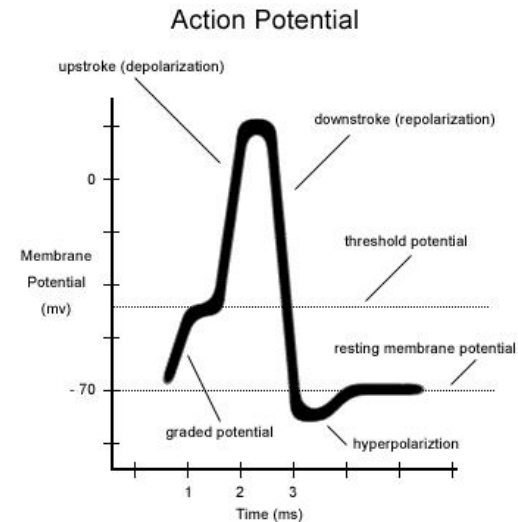
Dr. Rodolfo C. Raga Jr.
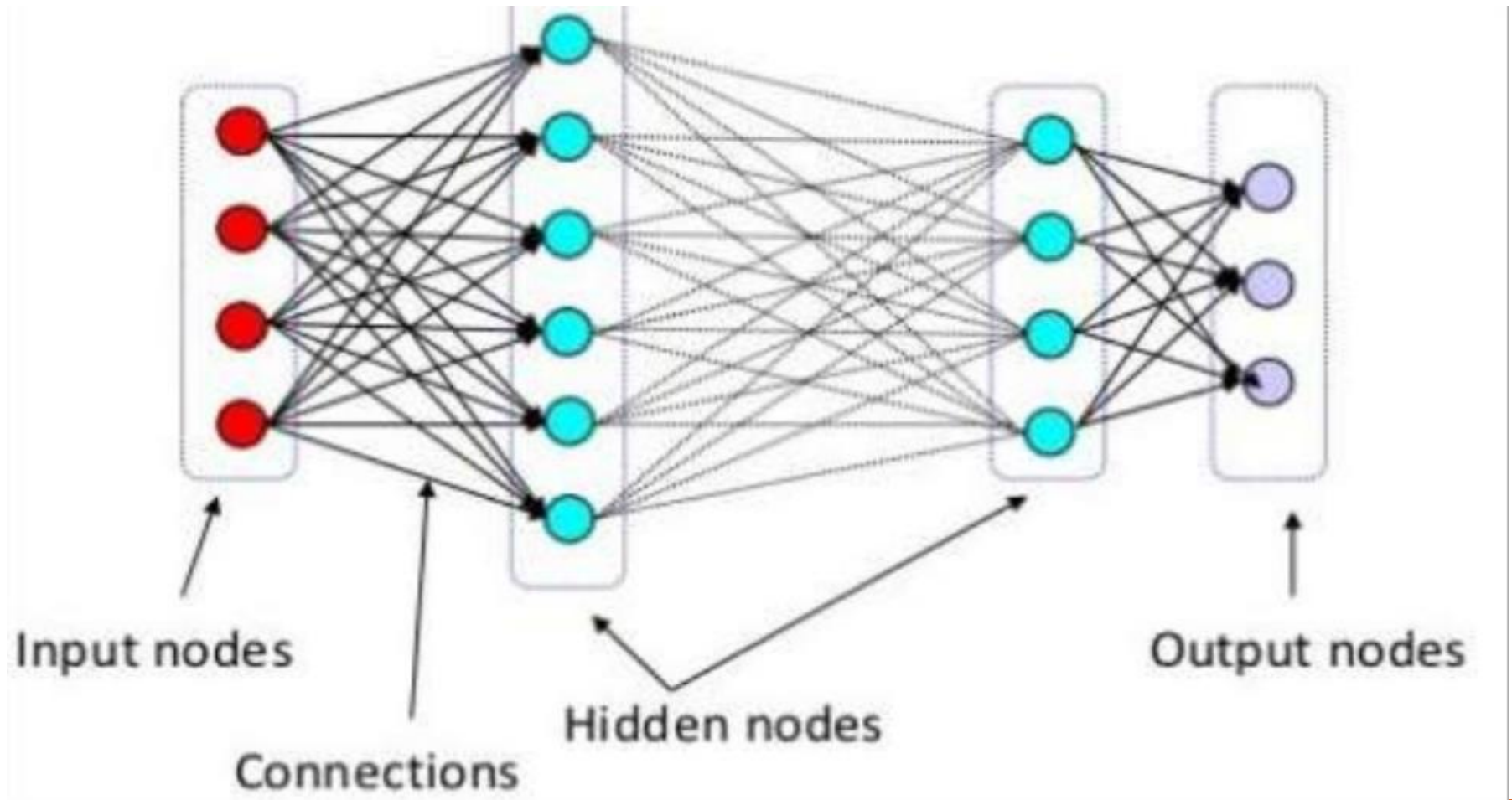NEUST Workshop – Dec 21-23, 2021
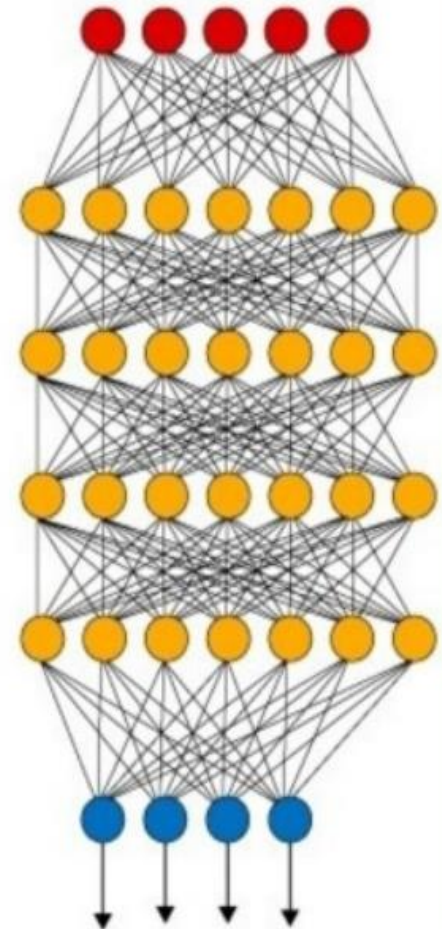
From neurons to ANNs

ANN is the only algorithm that simulates the functions of the human brain…

inspiration

$$y = \sigma\left(\sum_{i=1}^{N} w_i x_i + b\right)$$

Action Potential

activation function
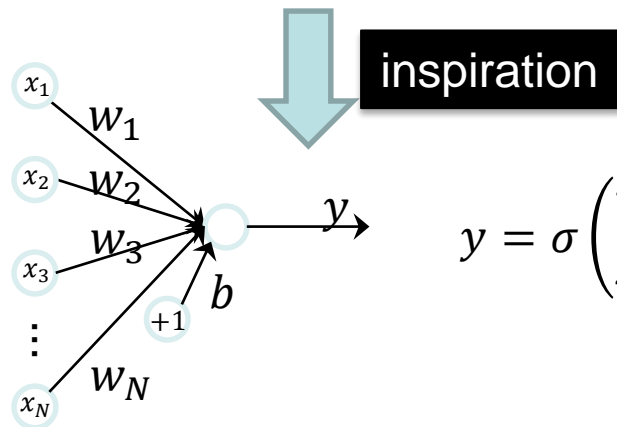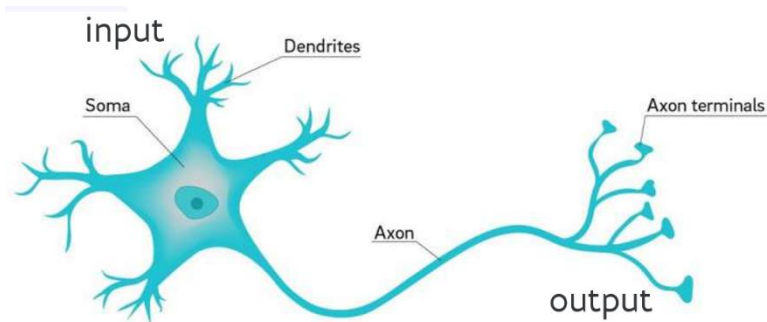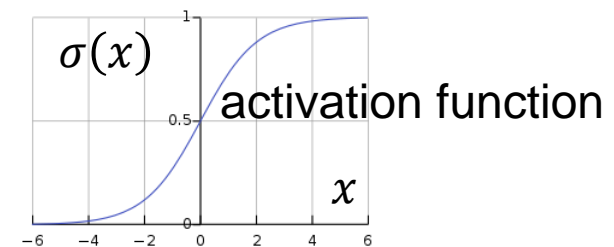
Deep Learning: Basic Architecture

Input nodes

Connections

Hidden nodes

Output nodes

Deep Learning: Basic Mechanism
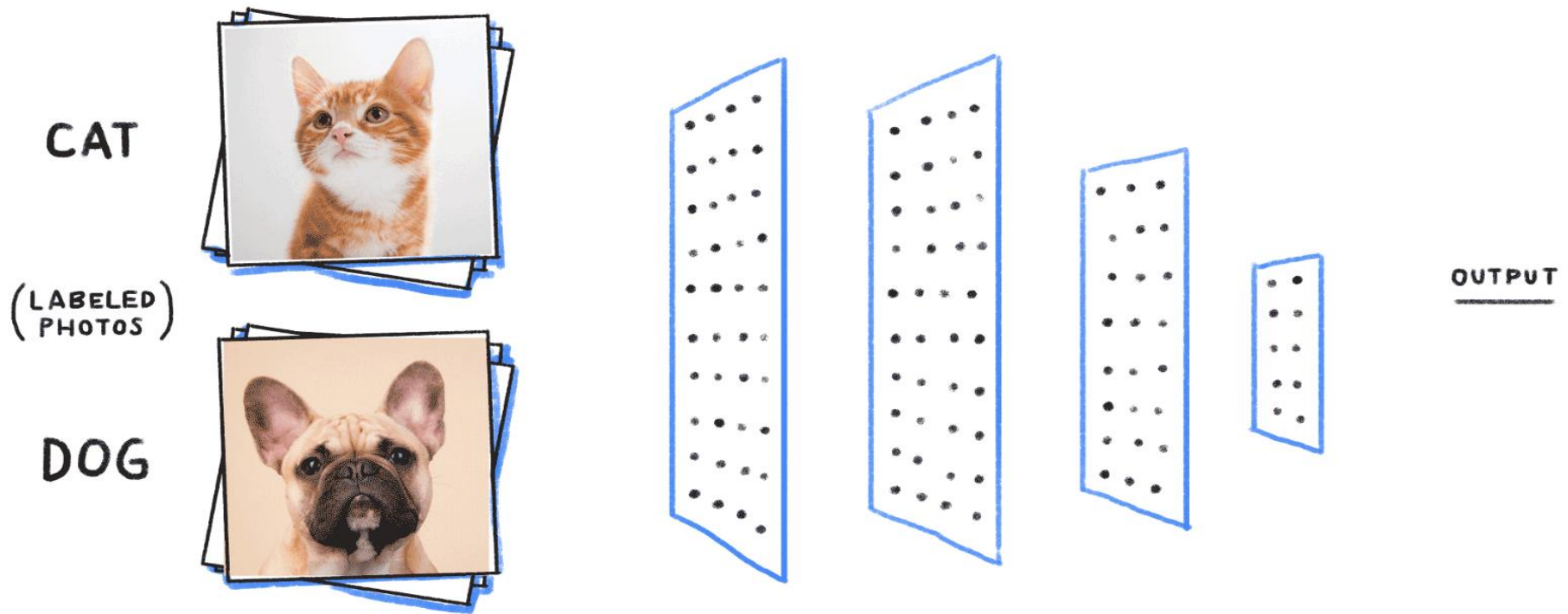
Deep Learning: How far can it go?

Image Detection Tasks

person, sheep, dog
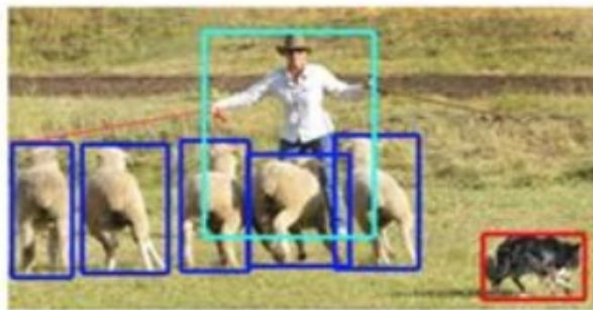
(a) classification   (b) detection   (c) segmentation

Detection task is harder than classification, but both are almost done.
And with better-than-human quality.

https://research.facebook.com/blog/learning-to-segment/

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Deep Learning: How far can it go?

Image recognition quality on ImageNet dataset
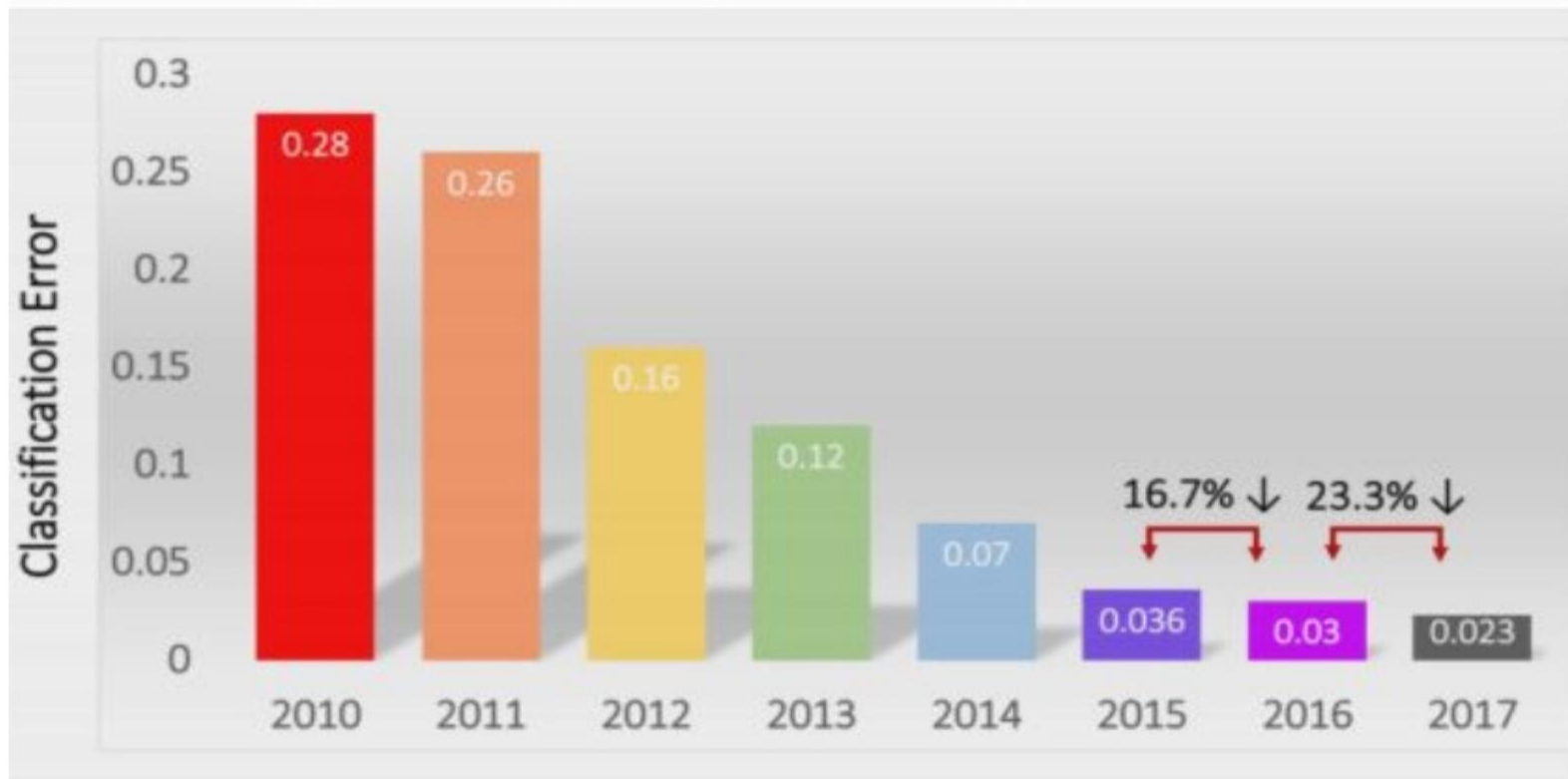
Human quality is estimated as ~5.1% error rate on this dataset (0.051)
From Lex Fridman slides: https://selfdrivingcars.mit.edu/

Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

Forbes

YOU TOLD US 31 NAMES WE NEED TO KNOW IN 2011

Forbes

TECH 2/19/2015 @ 1:06PM | 6,586 views

Microsoft's Deep Learning Project Outperforms Humans In Image Recognition

\+ Comment Now   \+ Follow Comments

Anthony Wing Kosner
Contributor

FOLLOW

TECH 12/29/2014 @ 11:37AM | 75,350 views

Tech 2015: Deep Learning And Machine Intelligence Will Eat The World

\+ Comment Now   \+ Follow Comments

Despite what Stephen Hawking or Elon Musk say, hostile Artificial Intelligence is not going to destroy the world anytime soon. What is
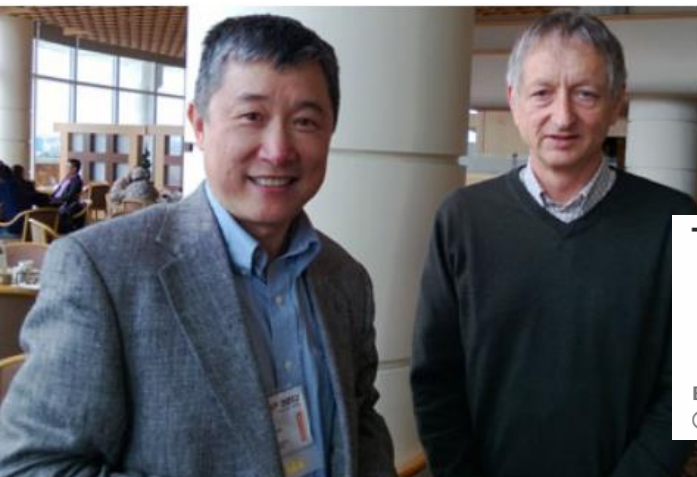
Dr. Rodolfo C. Raga Jr.
NEUST Workshop – Dec 21-23, 2021

In Academic World

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Affiliations | Corresponding author

"This joint paper from the major speech recognition laboratories was the first major industrial application of deep learning."

IEEE Signal Processing MAGAZINE

[VOLUME 29 NUMBER 6 NOVEMBER 2012]

LOUD AND CLEAR
FUNDAMENTAL TECHNOLOGIES IN MODERN SPEECH RECOGNITION

Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury

Deep Neural Networks for Acoustic Modeling in Speech Recognition

[The shared views of four research groups]

Application Demo

- Predicting Student Performance using Various ML algorithms

- Using Python for Image Classification

- Customized image classification using CNN