

DSC513 Introduction to Data Science

Dr. Ebin Deni Raj

Lead, Data Science Research Group

Indian Institute of Information Technology Kottayam

Managing your Expectation



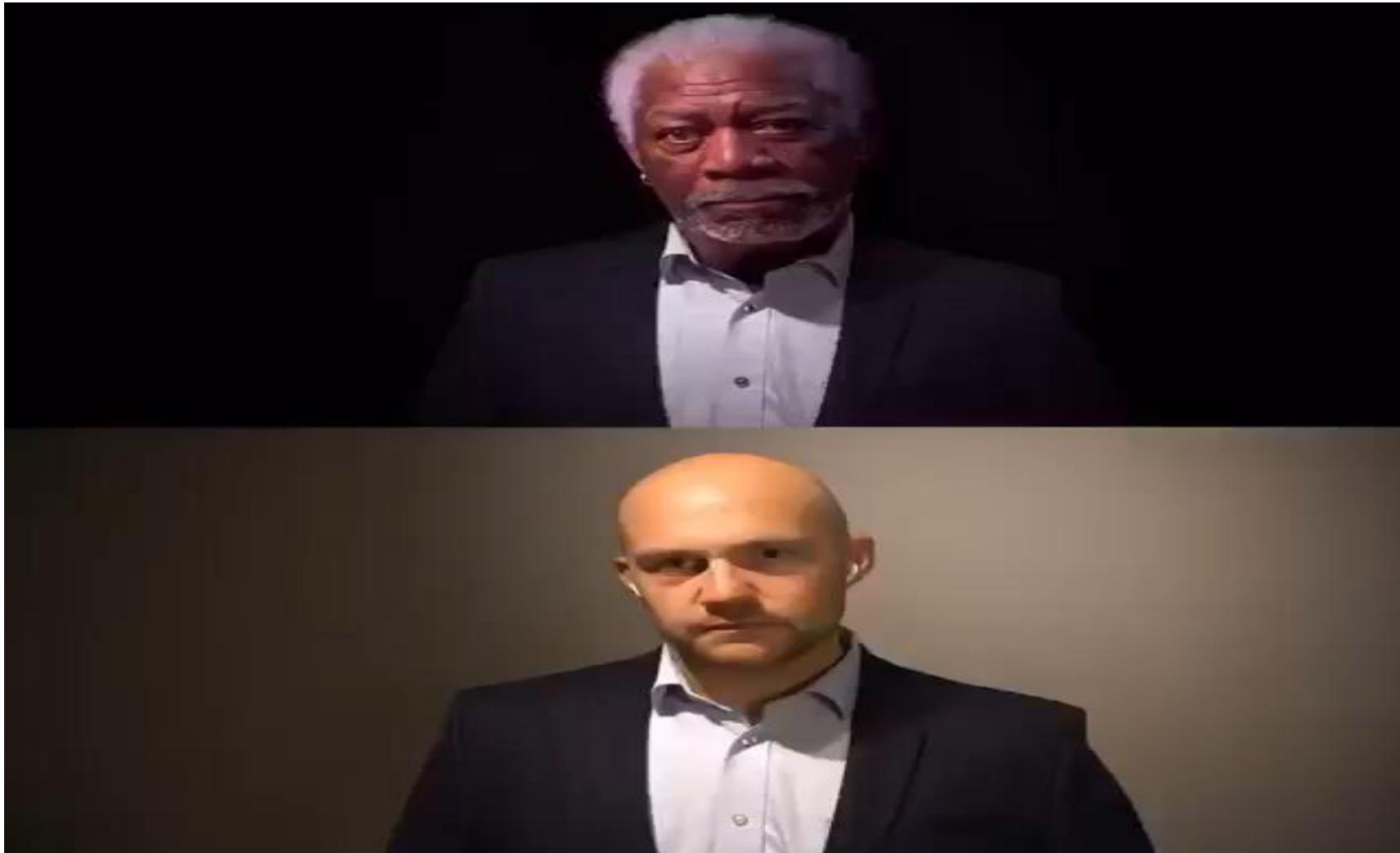
- Not an in depth research / technical session
- More of a “Big Picture” Type
- Will focus on the Industry needs of data science
- Will mention few things which our team is working on.

AI and Data Science in Action !!!



Source: MachineLearningmastery

This is not Morgan Freeman



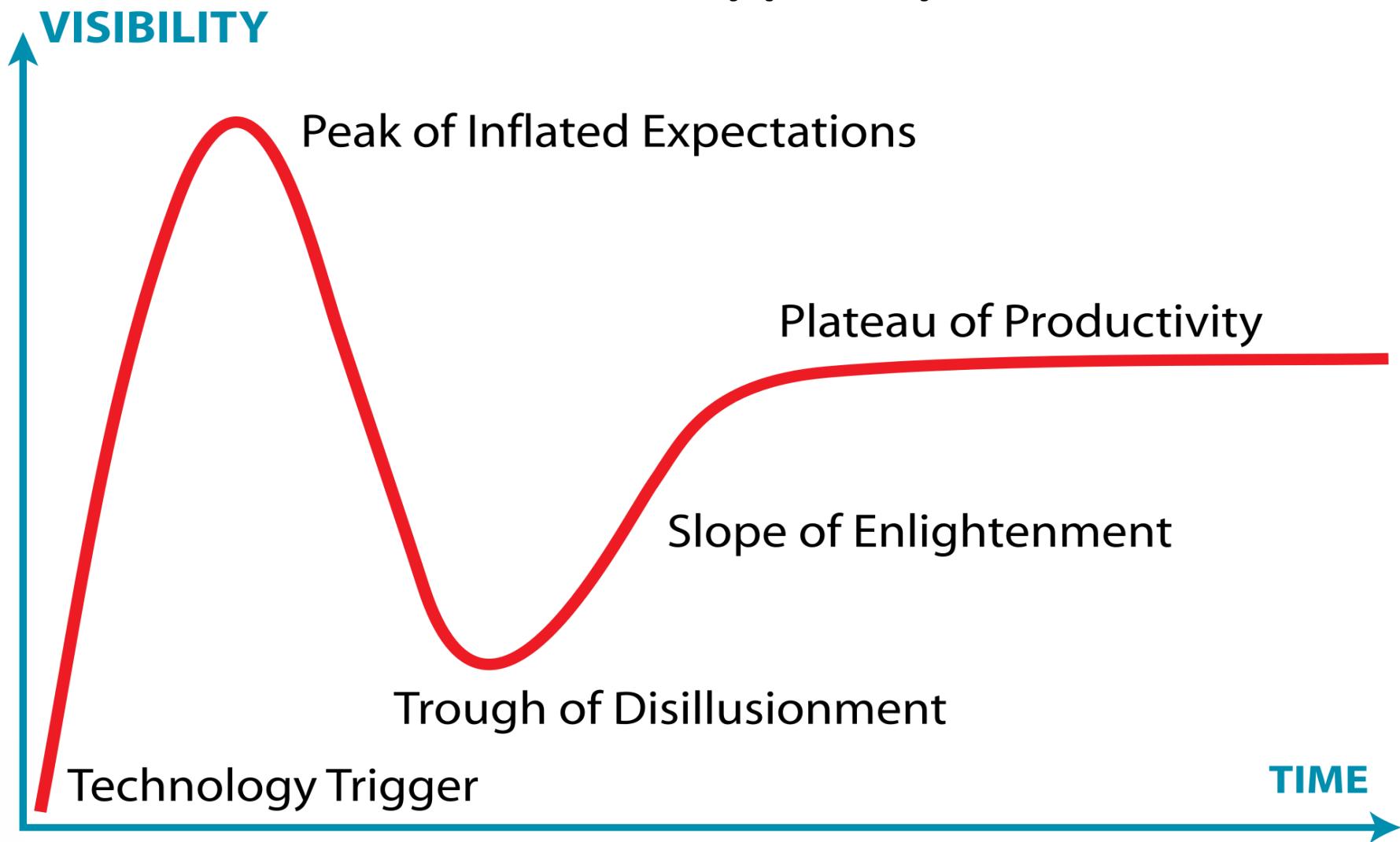
Looking back at History



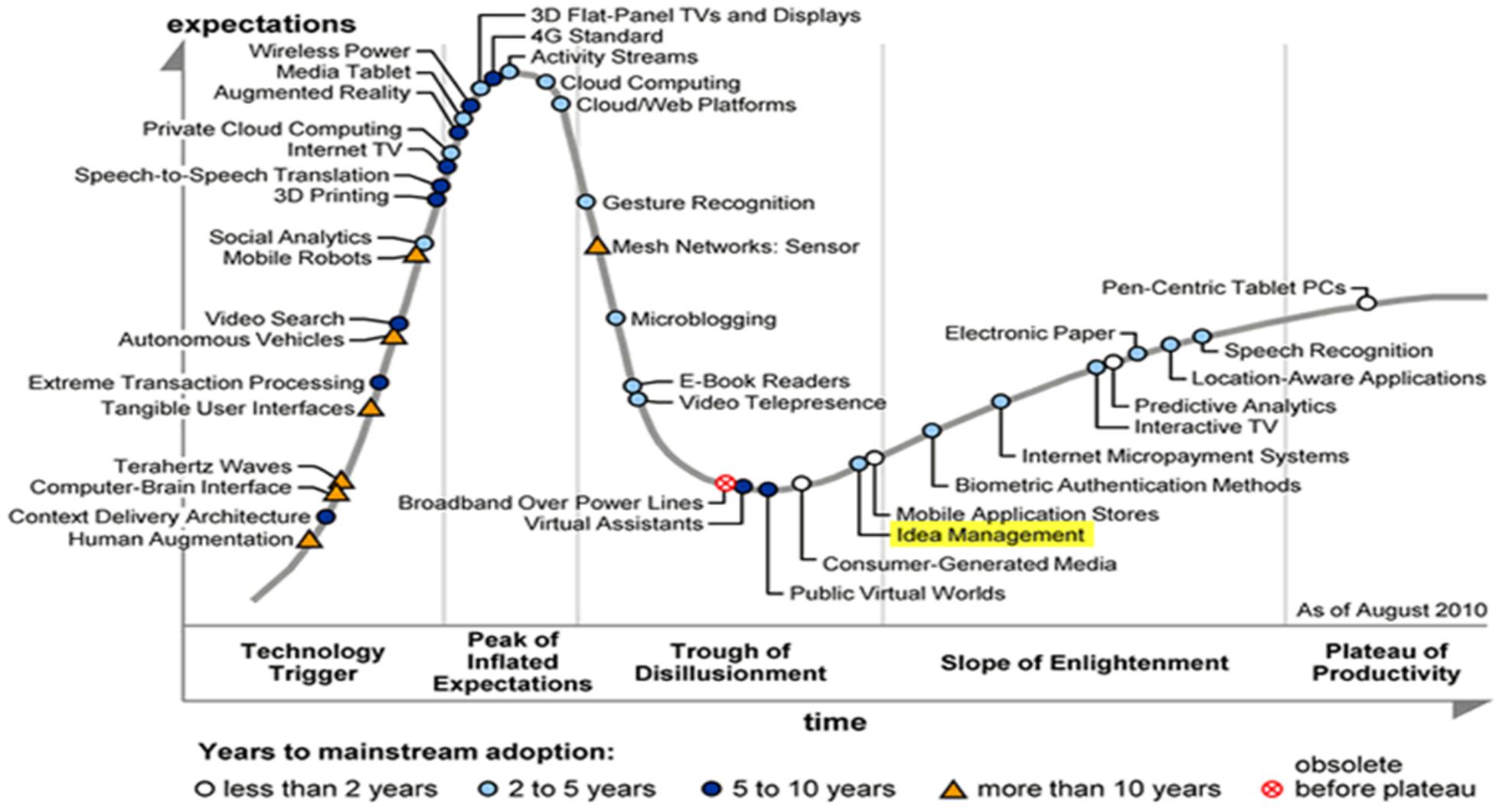
Data Analytics : Why?

- Why is it so popular?
- Practical and Necessary Skill for Business Careers
- Last two Decades → Computer Literacy
 - Word/Excel
 - Email/Web
- Next Two Decades → Analytics Literacy?
 - Excel too weak for analytics
 - How come each school is offering their own version?

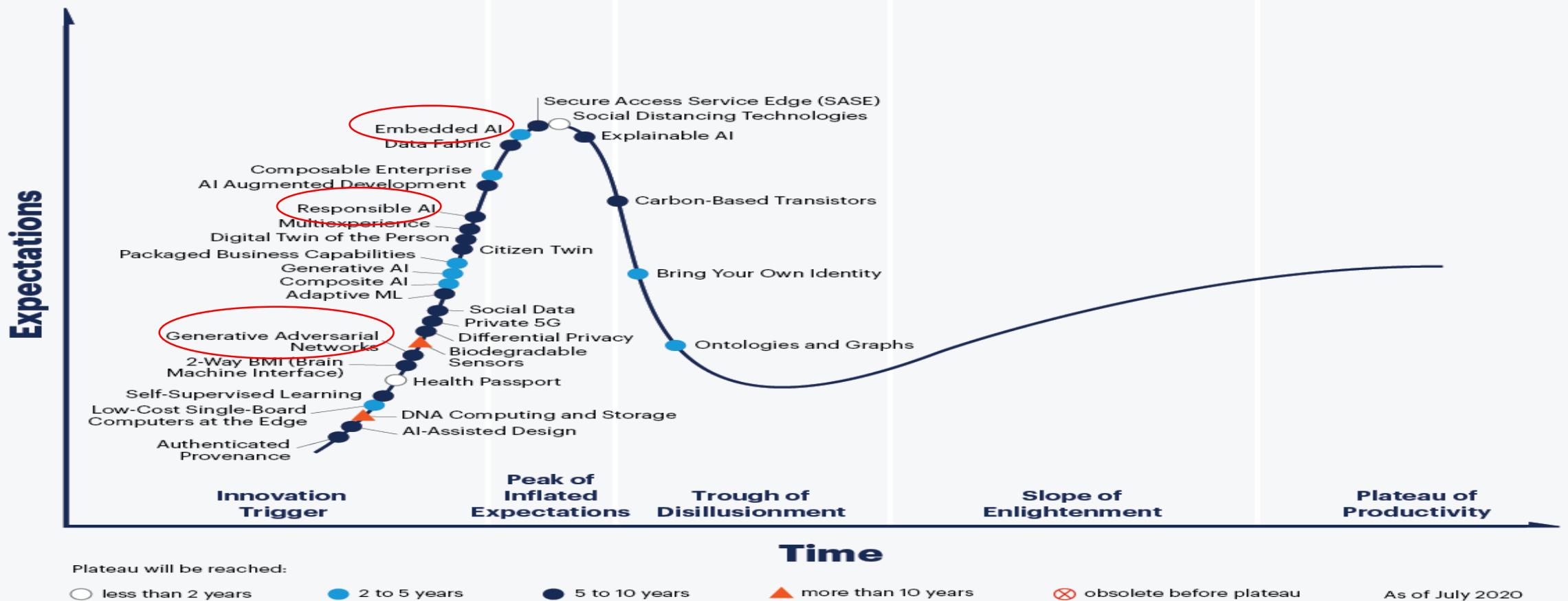
Gartner Hype cycle



Gartner Hype Cycle for Emerging Technologies 2010



Hype Cycle for Emerging Technologies, 2020



gartner.com/SmarterWithGartner

Source: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner®

Gartner Hype Cycle for Artificial Intelligence, 2019

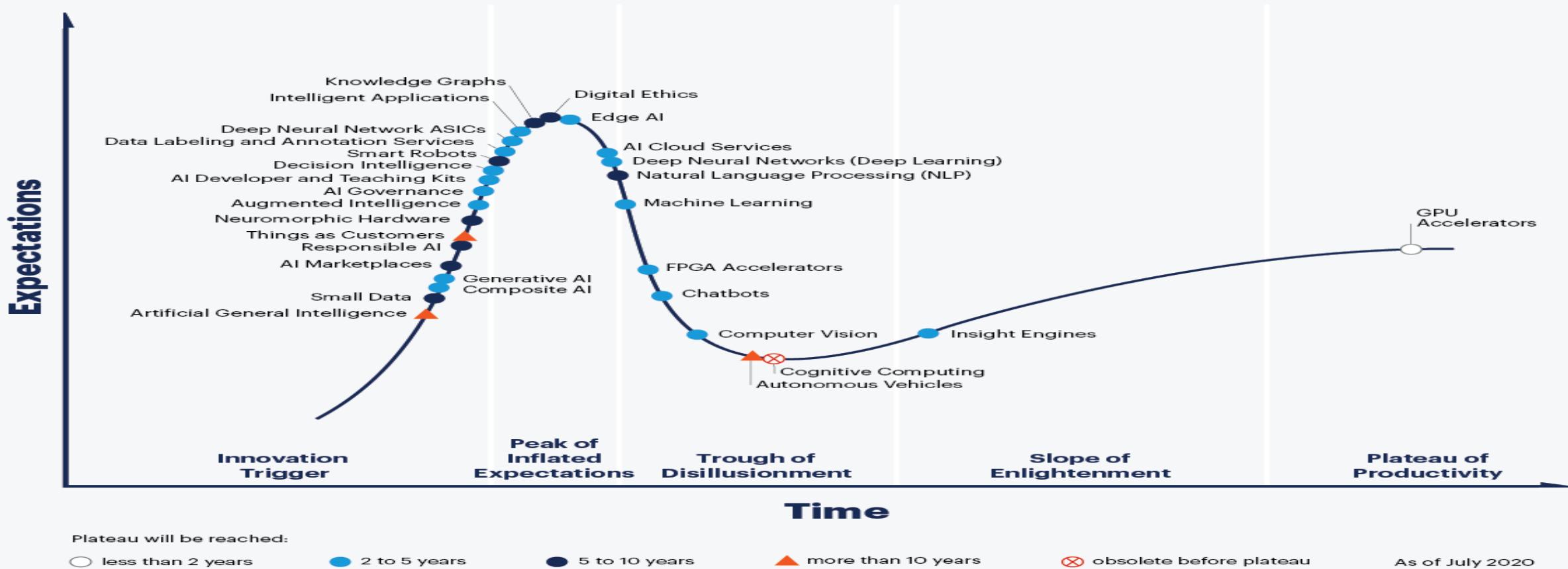


gartner.com/SmarterWithGartner

Source: Gartner
© 2019 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Hype Cycle for Artificial Intelligence, 2020

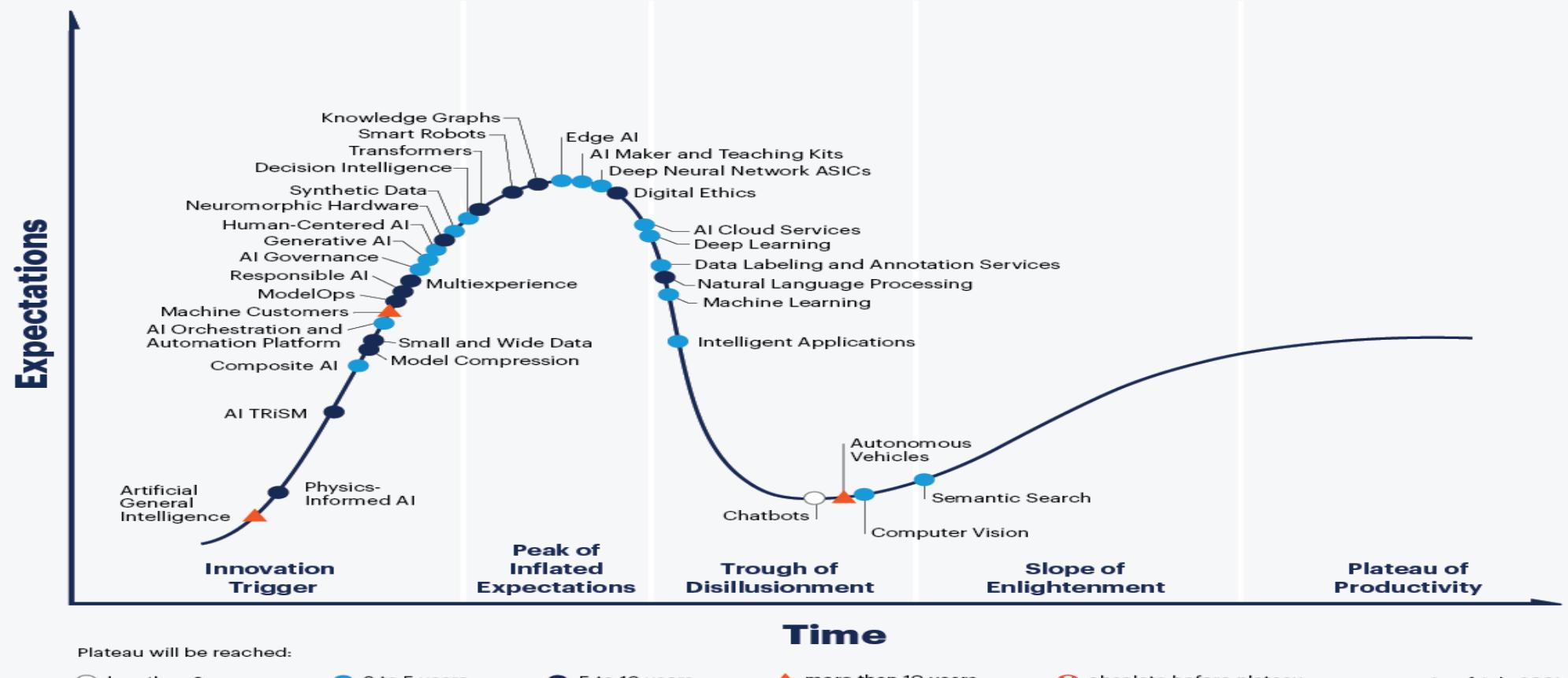


gartner.com/SmarterWithGartner

Source: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner

Hype Cycle for Artificial Intelligence, 2021



gartner.com

Source: Gartner
© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S. 1482644

Gartner®

Pause and Reflect

- Take a minute to think about Gartner Hype cycle and move your eyes away from the screen.

Break for exercise...

00:00

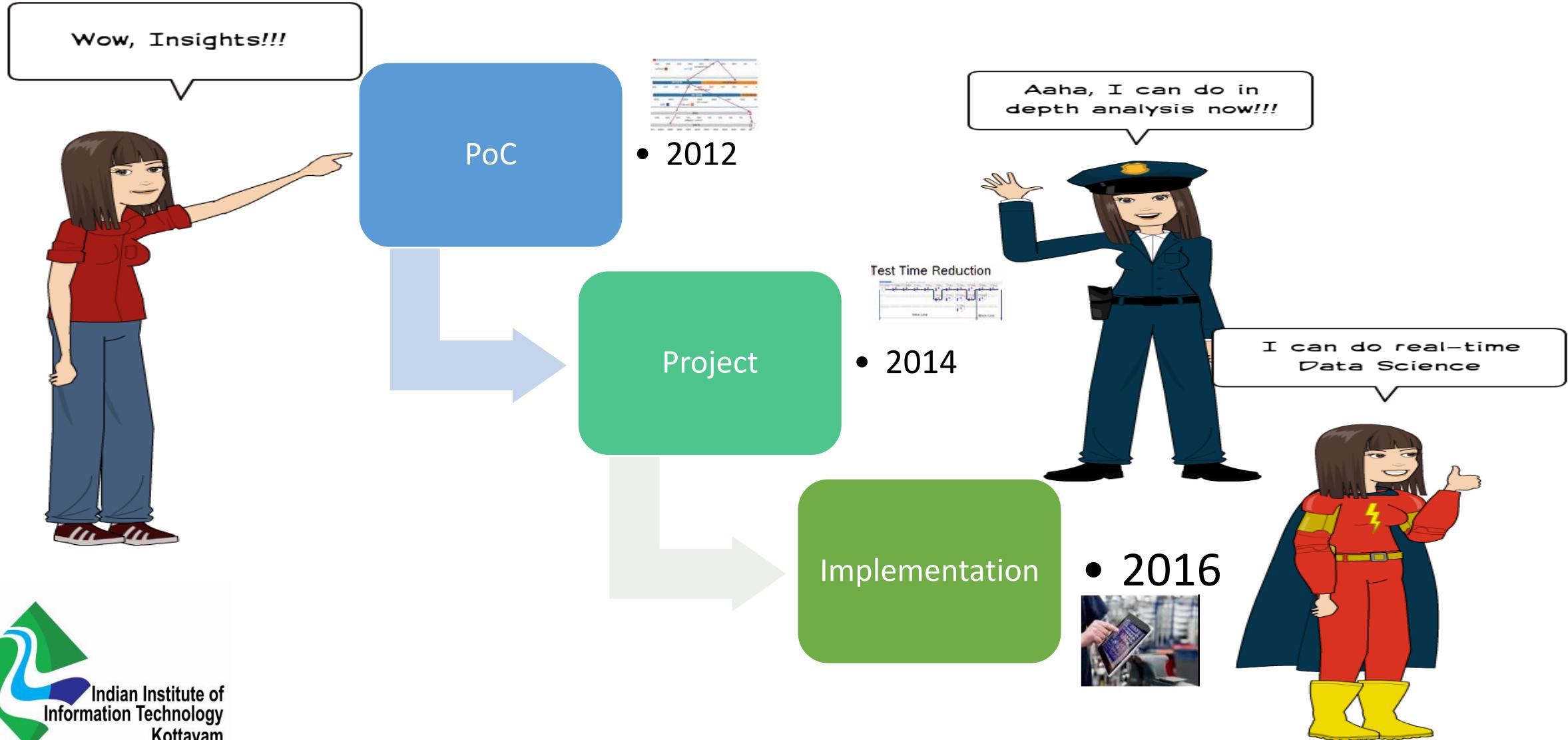


Question 1

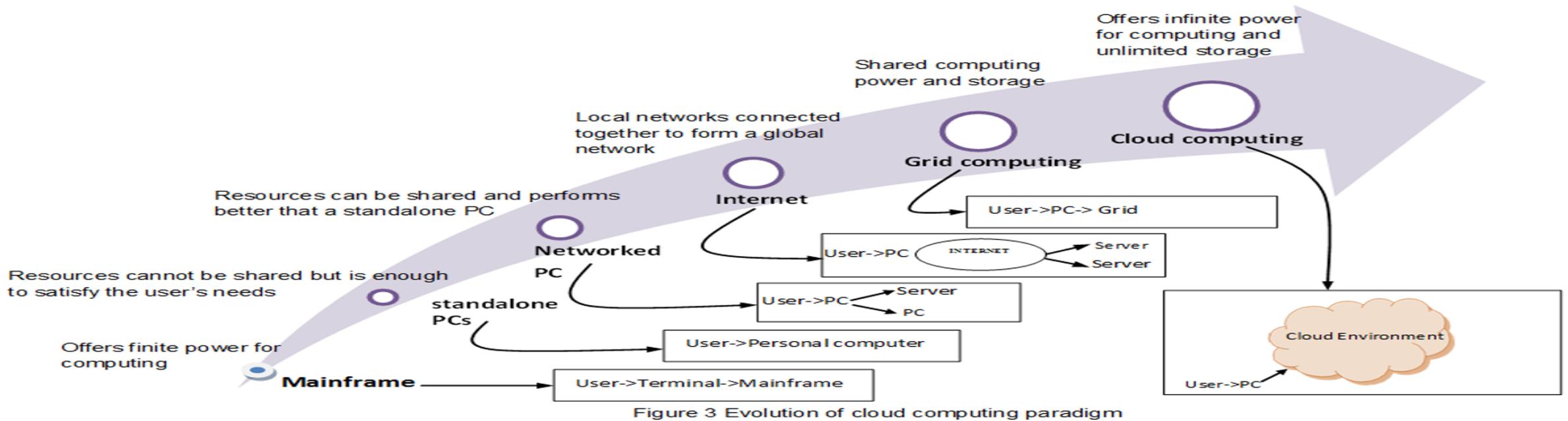
- The Plateau of Productivity is :
 - A. where you start seeing real world benefits of the technology
 - B. Media hype is the most
 - C. Where results begins to replace the original excitement about potential value
 - D. industry interest is at the peak in a technology innovation



Where does Data Science stand Today?



Evolution of Computing



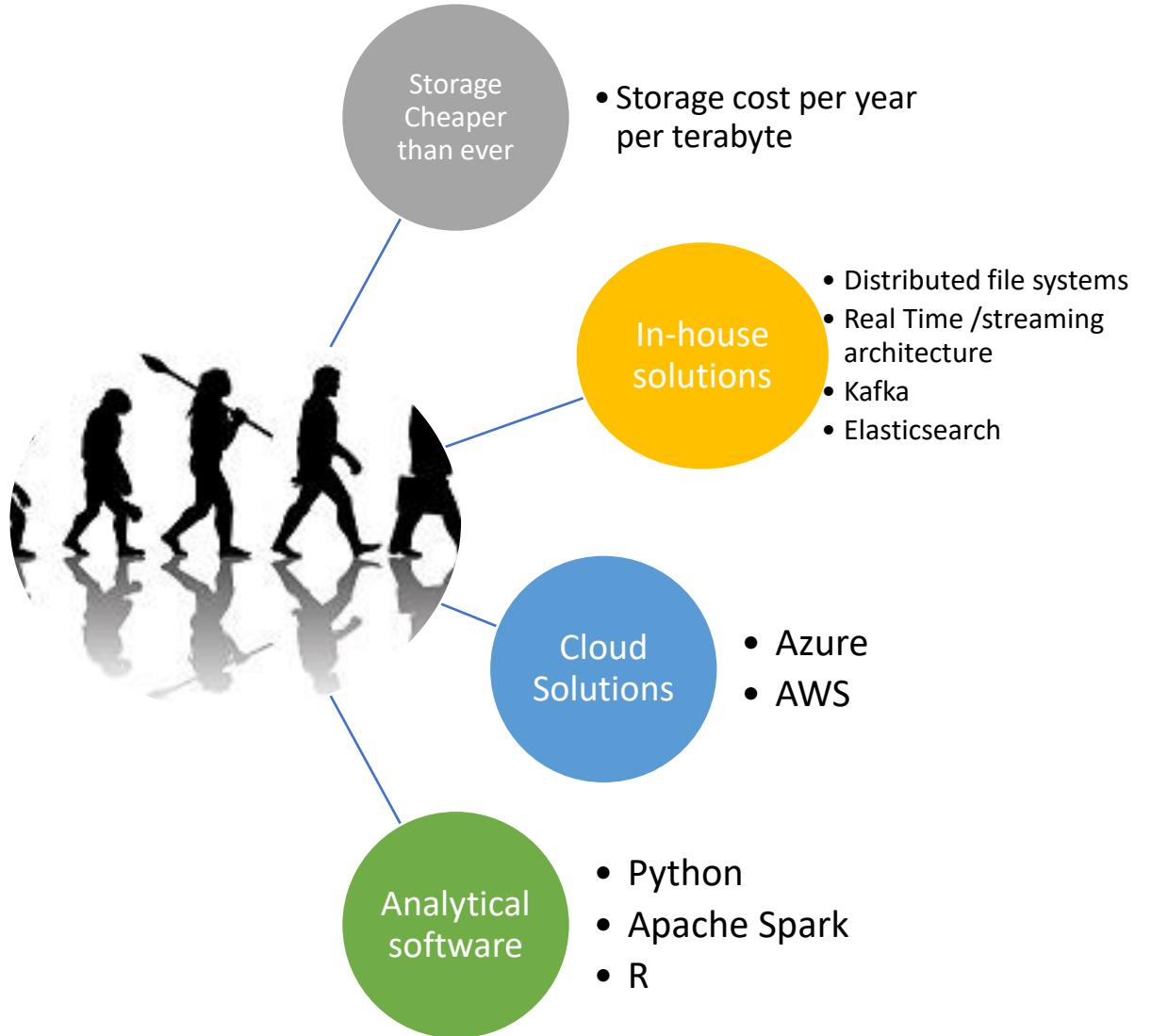
Source: Cloud Technology: Concepts, Methodologies, Tools, and Applications, 2354-2372, Author: Ebin Deni Raj



© Go Physics

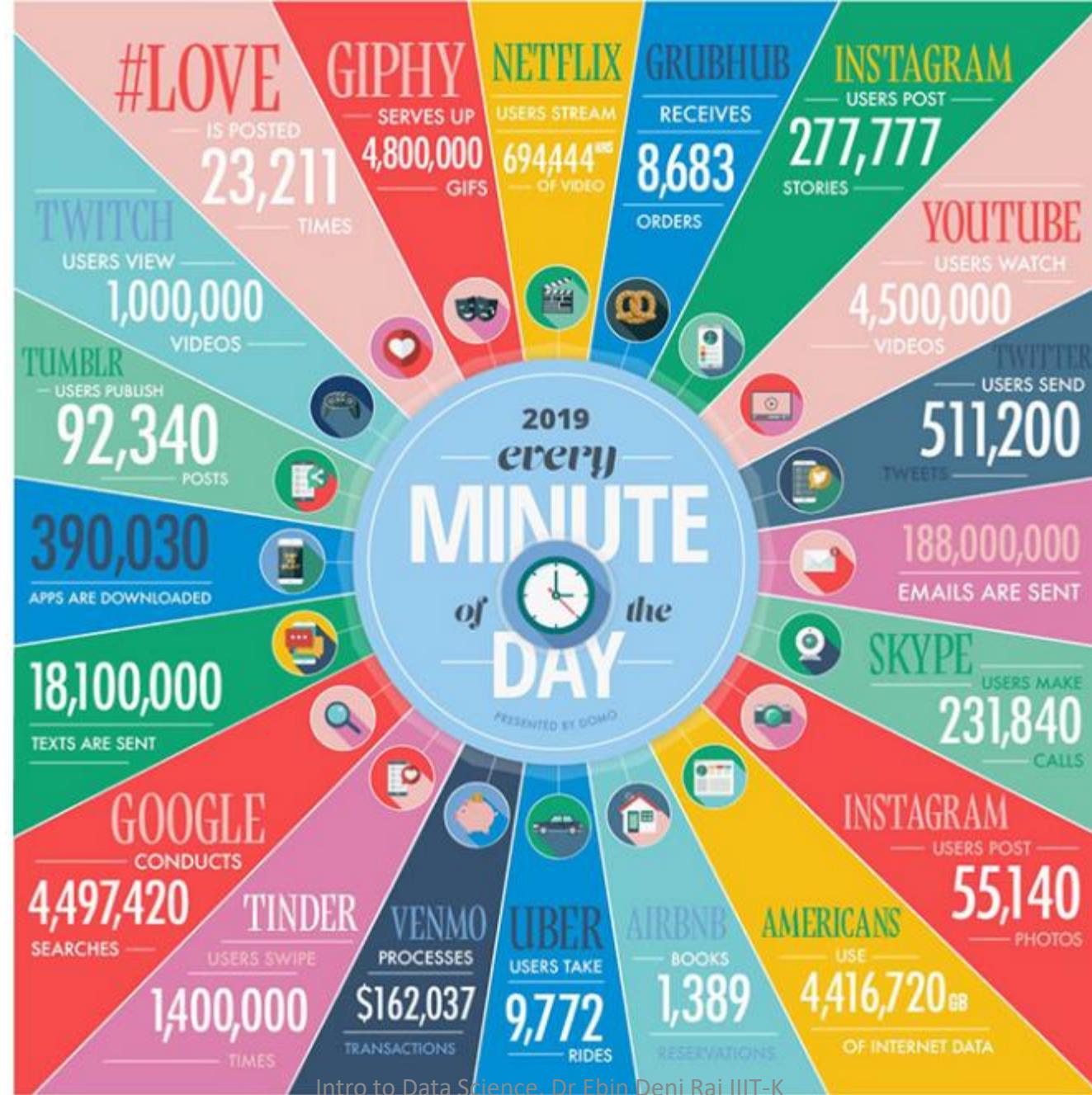


Technology Evolutions





In 2019



In 2020



In 2021





Data: the new currency?



“If you are not paying for it,
You are the product.”

– Some really smart guy on the internet –

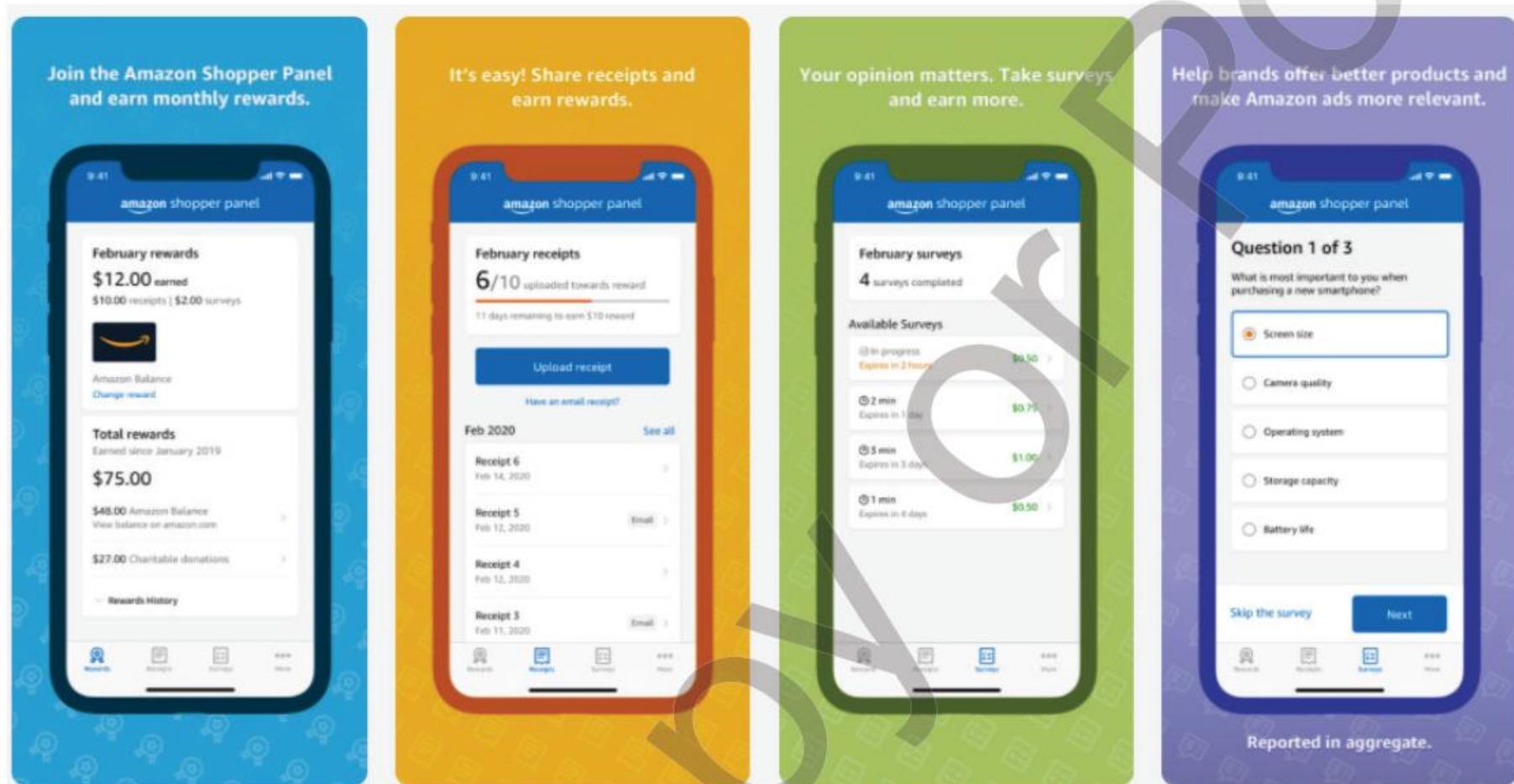
Facebook Rebuked for Failing to Disclose Data-Sharing Deals



A sign outside Facebook's headquarters in Menlo Park, Calif. The social network faced
sharp criticism after a New York Times article revealed details about its extensive data

- Source: <https://www.nytimes.com › technology › facebook-data-privacy-criticism> Dated Dec19, 2018

Amazon Shopper: Paying customers for their Data



Source: Amazon.com (2020), <https://panel.amazon.com/>, accessed December 2020.

Intro to Data Science, Dr Ebin Deni Raj IIIT-K

THE HINDU BusinessLine

← Ads by Google Stop seeing this ad Why this ad? ⓘ

SENSEX 61,223.03 ▼ -12.27 [-0.02%] Tata-Steel 1213.35 ▼ -7.25 [-0.59%] TCS 3969.25 ▲ NIFTY 18,255.75 ▼ -2.05 [-0.01%] 58% HCL-Technologies 1337.2

≡ HOME NEWS MARKETS COMPANIES INFO-TECH OPINION SPECIALS PORTFOLIO PREMIUM MORE ▾ Search here

LATEST NEWS | BOOKS | BL EXPLAINER | ECONOMY | AGRI BUSINESS | BL ON CAMPUS | MONEY & BANKING | DATA FOCUS | CARTOONS | RESOURCES

Info-tech

84% Indian consumers willing to pay more to do business with organisations committed to protecting data privacy: Report

Hemai Sheth | Mumbai, March 14 | Updated On: Mar 14, 2021





“Personal data is the new oil of the internet and the new currency of the digital world.”

— Meglena Kuneva
European Consumer Commissioner

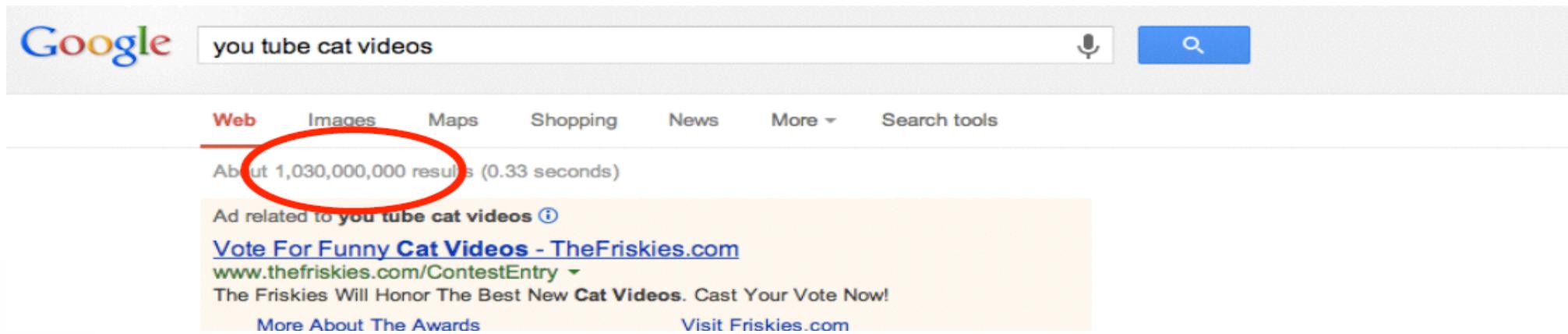
Deluge!!!



Big Data

“Between the dawn of civilization and 2003, we only created five exabytes of information; now we’re creating that amount every two days.”

Eric Schmidt, Google (and others)



A screenshot of a Google search results page. The search query "you tube cat videos" is entered in the search bar. Below the search bar, the "Web" tab is selected, while "Images" is highlighted with a red circle. The search results section displays the following information:

- "About 1,030,000,000 results (0.33 seconds)"
- An advertisement for "Ad related to you tube cat videos" from "Vote For Funny Cat Videos - TheFriskies.com". The ad includes the URL www.thefriskies.com/ContestEntry and the text "The Friskies Will Honor The Best New Cat Videos. Cast Your Vote Now!".
- Links for "More About The Awards" and "Visit Friskies.com".

Too big to handle



Mega Trends in Data Landscape



- Explosion of Data
- Exponential Growth in Data volume in many fields and expected to continue



- Analytical and Reporting Requirements
- Demand for more sophistication from business, customers and regulators



- Need for Real Time Speed
- Almost real time requirement for processed output => Data Output



- Systems Innovations
- Drop in Technology Cost. Innovative platforms for storage volume, velocity => Big Data

Huge Demand for Data Driven Insights

Service Improvements



Process Improvements



New Product Offerings



Novel Business Models



Hospitality



Retail



Entertainment



Apparel



Jiuzhaigou National Park project



More Examples



The AI-Powered Online Fitting Room: VOGUE



An AI Software Able To Detect and Count Plastic Waste in the Ocean

AI Application in



E-commerce



Education



Lifestyle



Navigation



Robotics



Human Resource



Health care



Agriculture



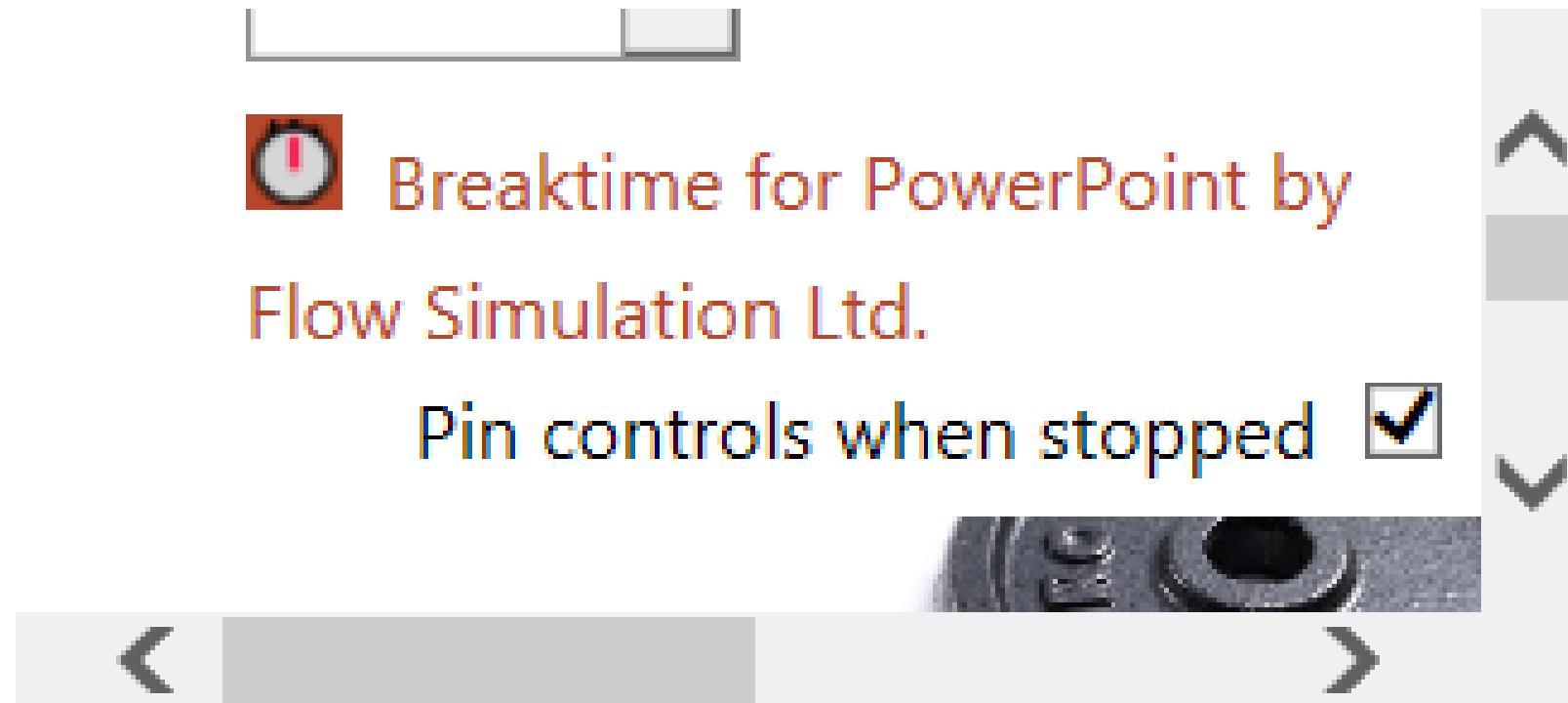
Marketing



Chatbots

Pause and Reflect

- What Huge data is coming your way? What can you do with it?
- Take a minute to think about that and move your eyes away from the screen.

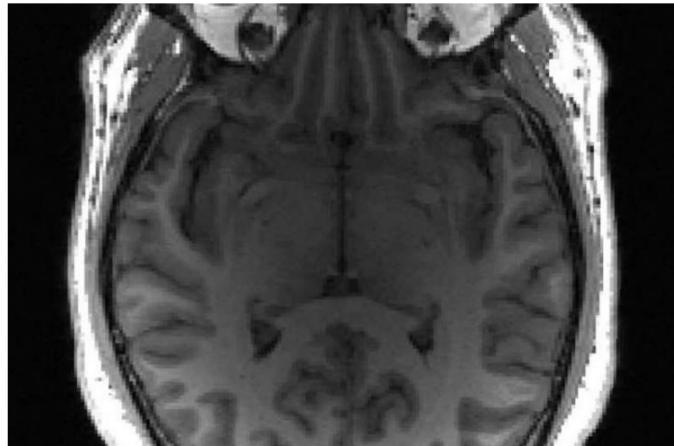


Application in Variety of Disciplines

① AUGUST 12, 2021

AI could detect dementia years before symptoms appear

by University of Cambridge



MRI brain scan of healthy volunteer. Credit: Timothy Rittman

Dementias are characterized by the build-up of different types of protein in the brain, which damages brain tissue and leads to cognitive decline. In the case of Alzheimer's disease, these proteins include beta-amyloid, which forms 'plaques,' clumping together between neurons and affecting their function, and tau, which

Can technology help authors write a book?

By Bernd Debusmann Jr
Business reporter

① 5 days ago



Writing a book is difficult, but could using technology help the process? GETTY IMAGES

Celebrated American author Mark Twain was very dismissive of people who think it is possible for someone to learn how to write a novel.

"A man who is not born with the novel-writing gift has a troublesome time of it when he tries to build a novel," he said. "He has no clear idea of his story. In fact, he has no story."

British writer Stephen Fry puts it another way. He says that successful authors are those who know just how difficult it is to write a book.

Every year around the world a whopping 2.2 million books are published, according to the United Nations Educational, Scientific and Cultural Organization (Unesco), which monitors the number. The figure includes both fiction and non-fiction titles.

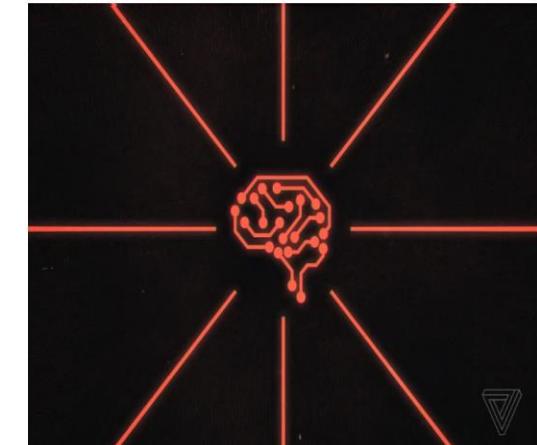
TECH ARTIFICIAL INTELLIGENCE

OpenAI can translate English into code with its new machine learning software Codex

'We see this as a tool to multiply programmers'

By James Vincent | Aug 10, 2021, 1:00pm EDT

f t SHARE



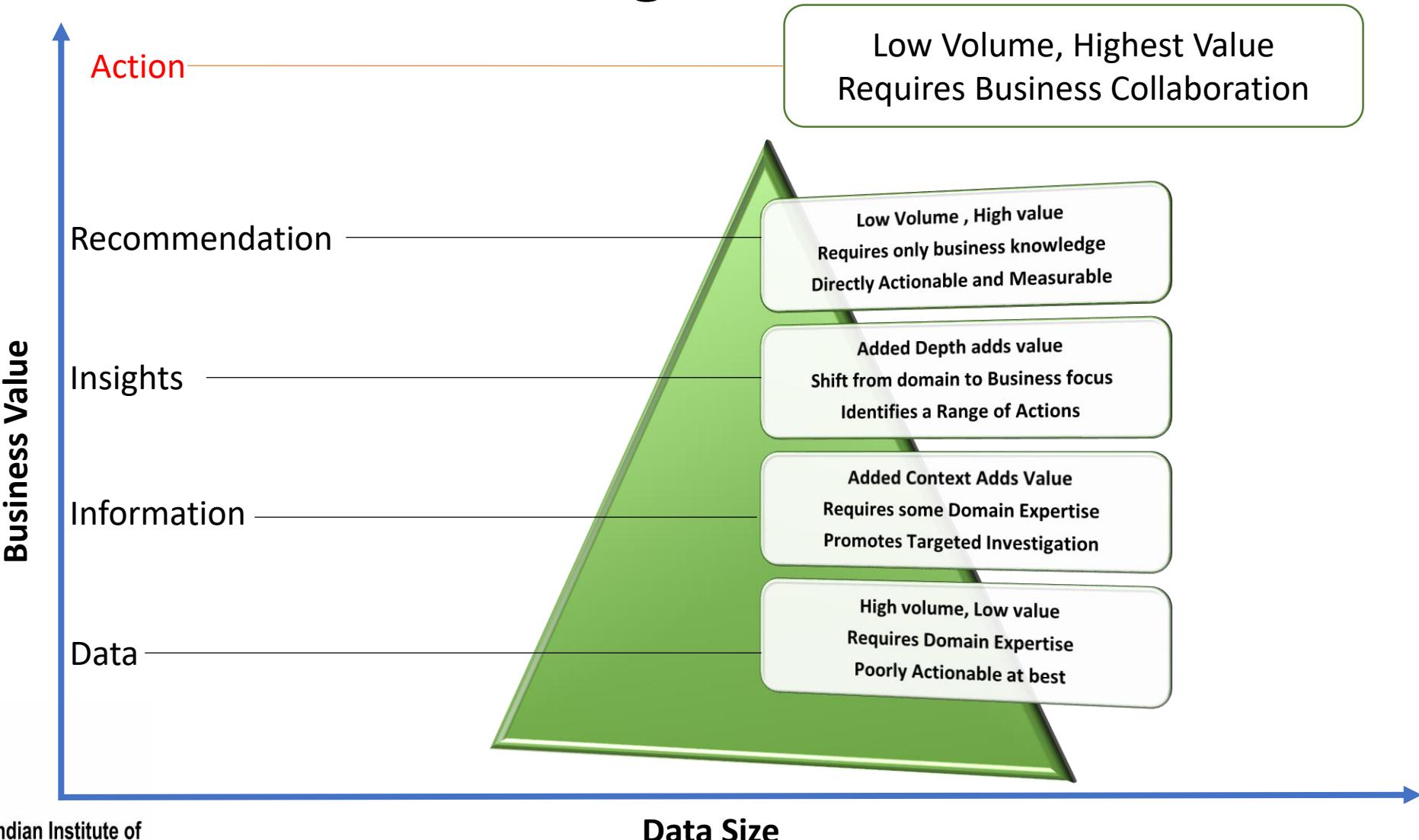
Ad closed by Google

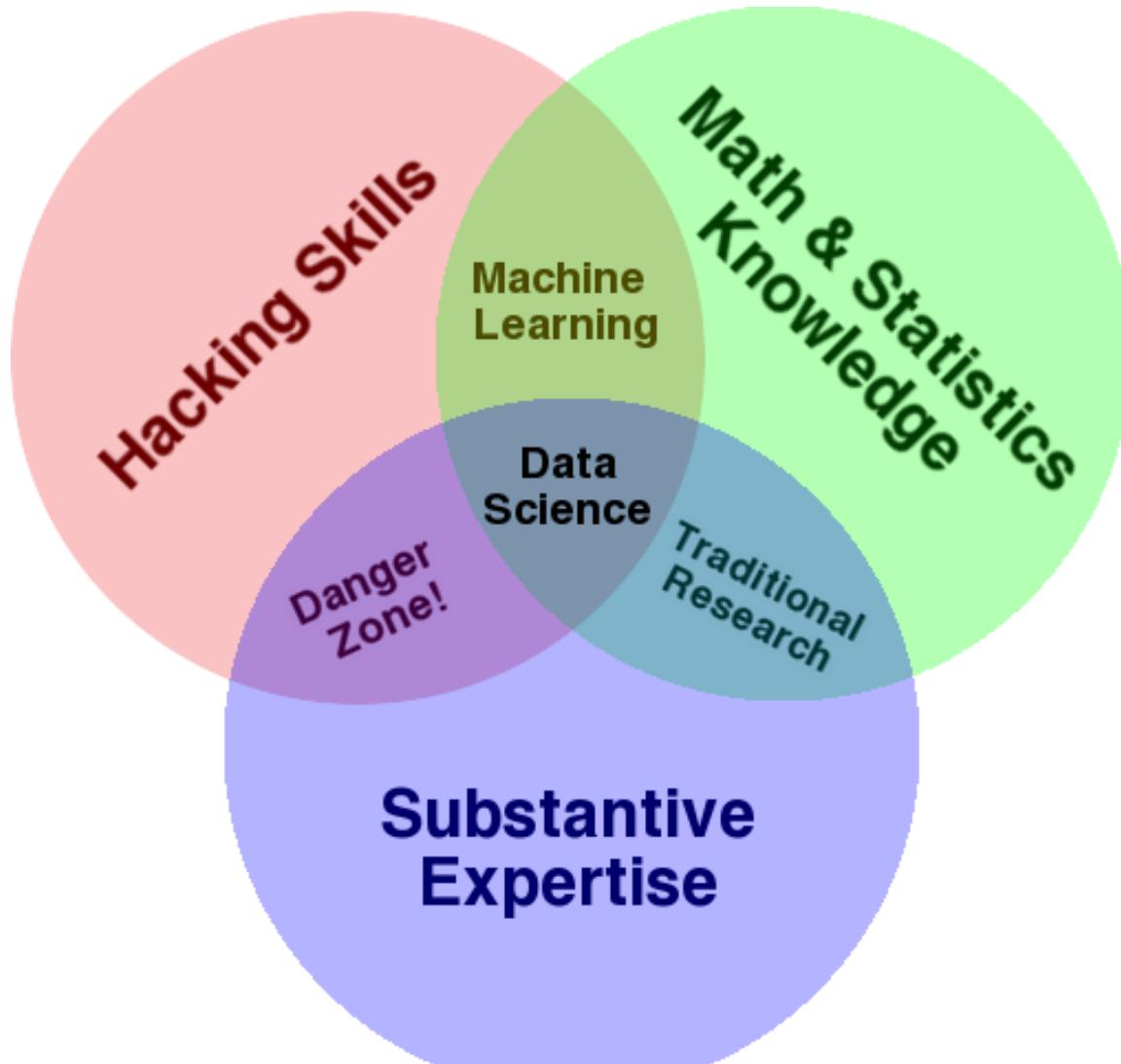
Question 2

- What has made the sudden boom in Data Science and allied areas?
 - A. Democratization of Internet
 - B. Smart devices usage
 - C. Computing got better
 - D. All of the above

Data Analytics : What?

Turning Data into Value

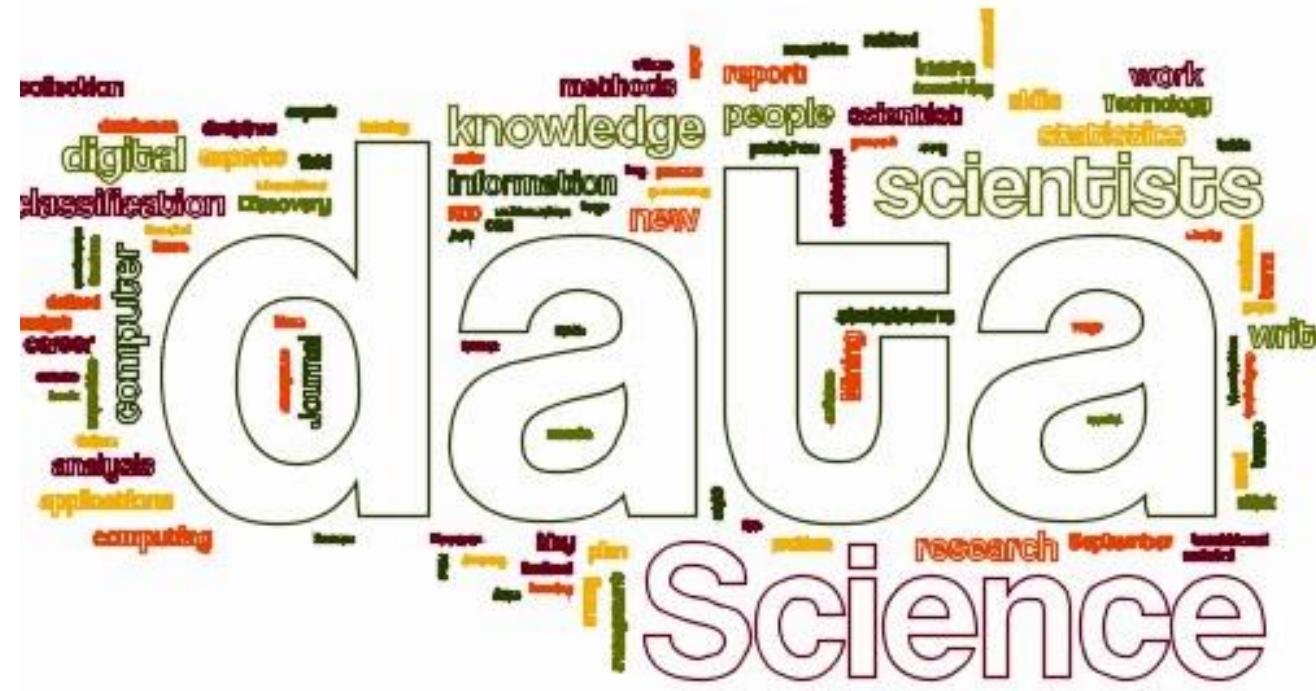




Drew Conway's Venn diagram of data science

[Source: Doing Data Science: Straight Talk from the Frontline]

Data Science Programming

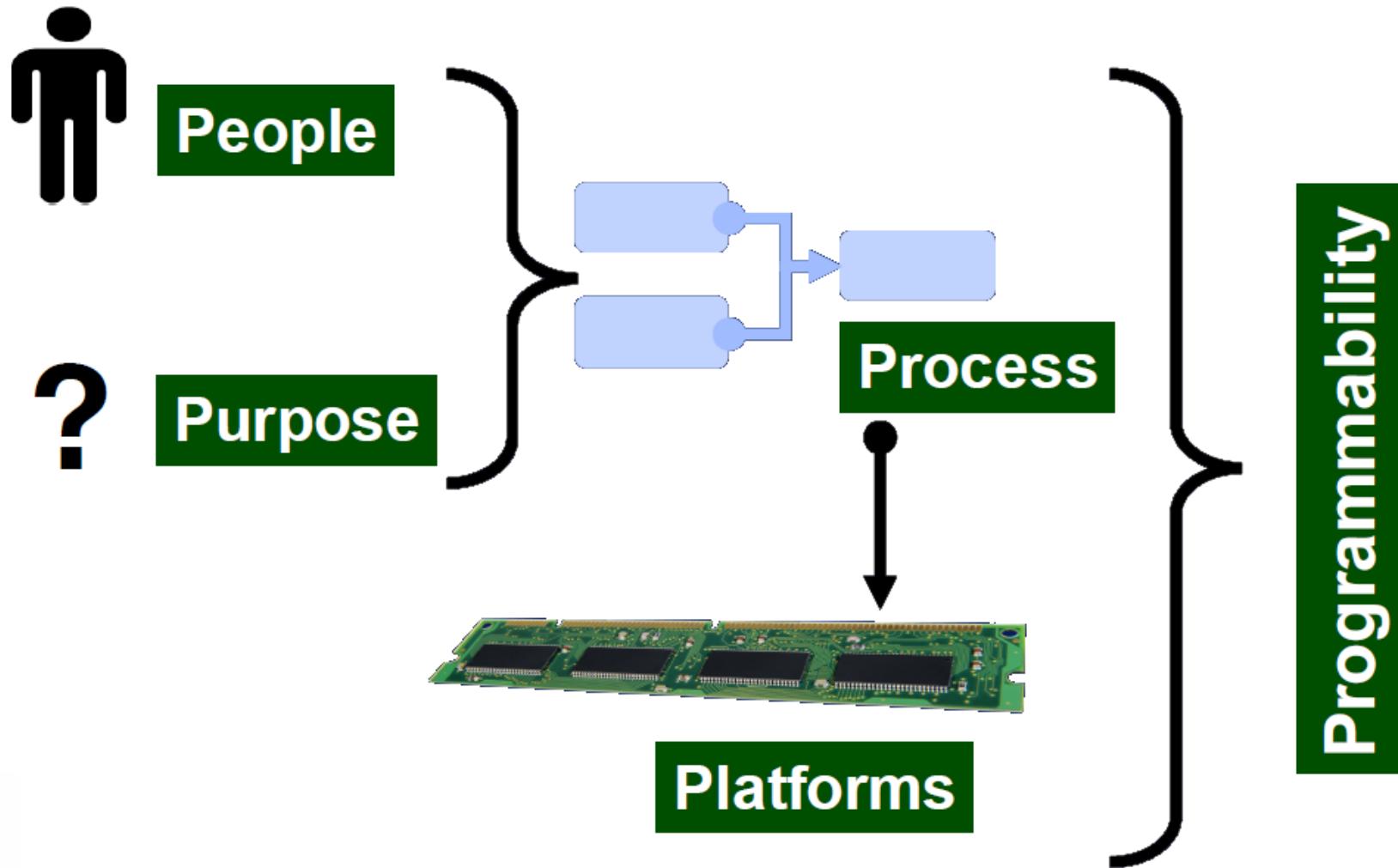


Question 03

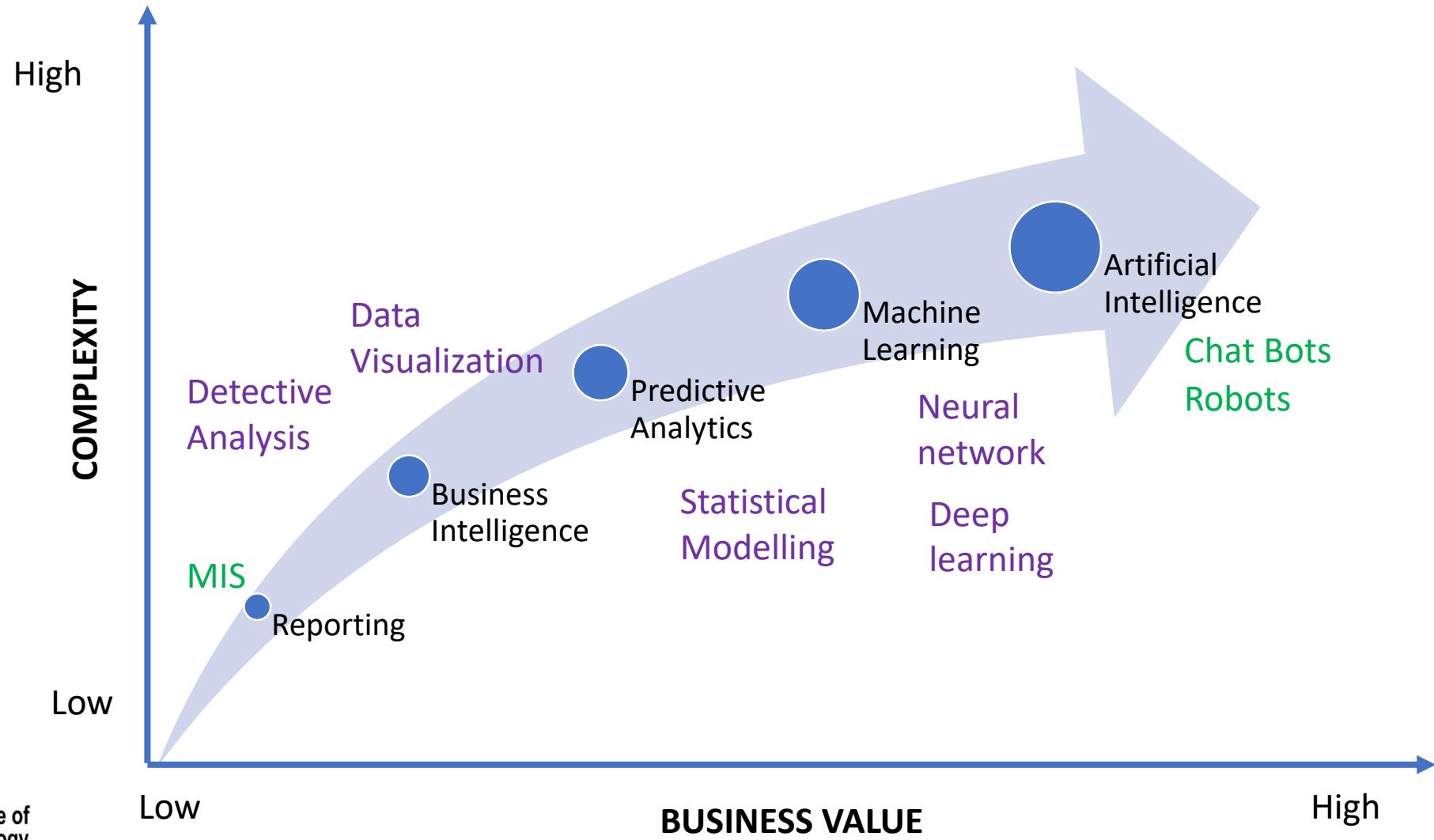
- As per Drew Conning's venn diagram A person with **Maths , Statistical Knowledge + substantive Expertise** which of the following skill :
 - A. Hacking skill
 - B. Machine Learning
 - C. Traditional research
 - D. Data Science



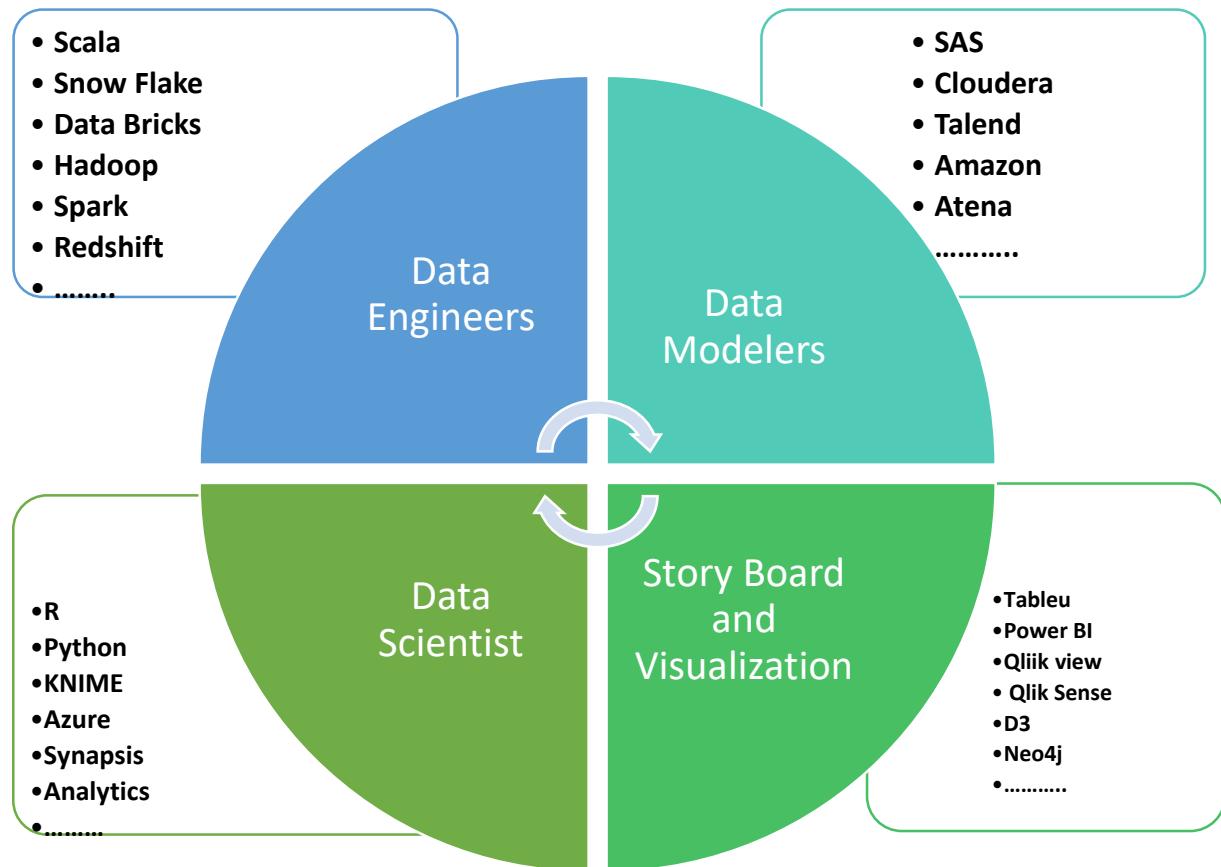
5 Ps of Data Science



Spectrum of Data Science



Breadth of Analytic Skill Base



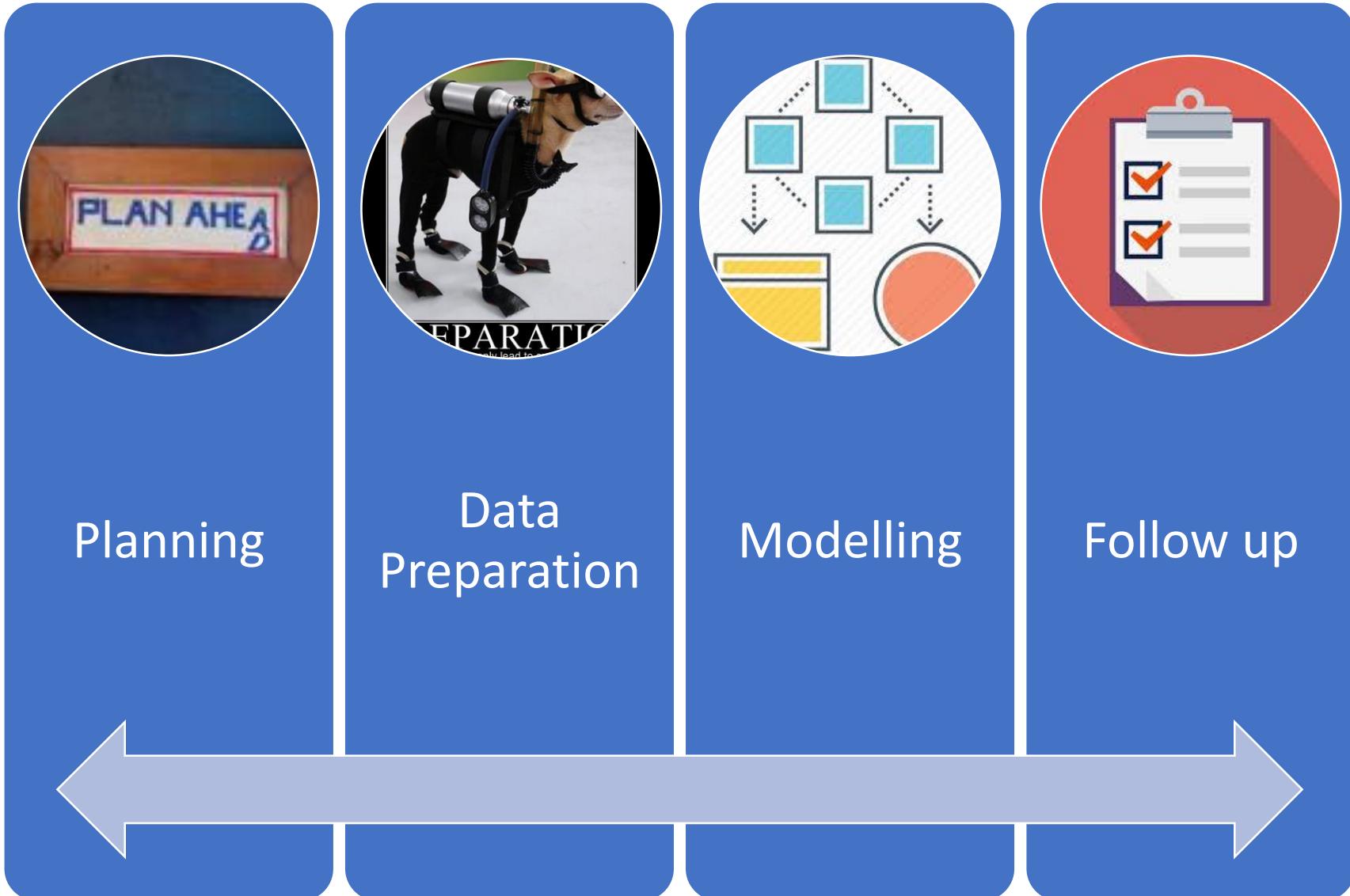
Question 04

- Which of the following is Open source Technology/tool?
 - A. SPSS
 - B. SAS
 - C. Apache Spark
 - D. Matlab

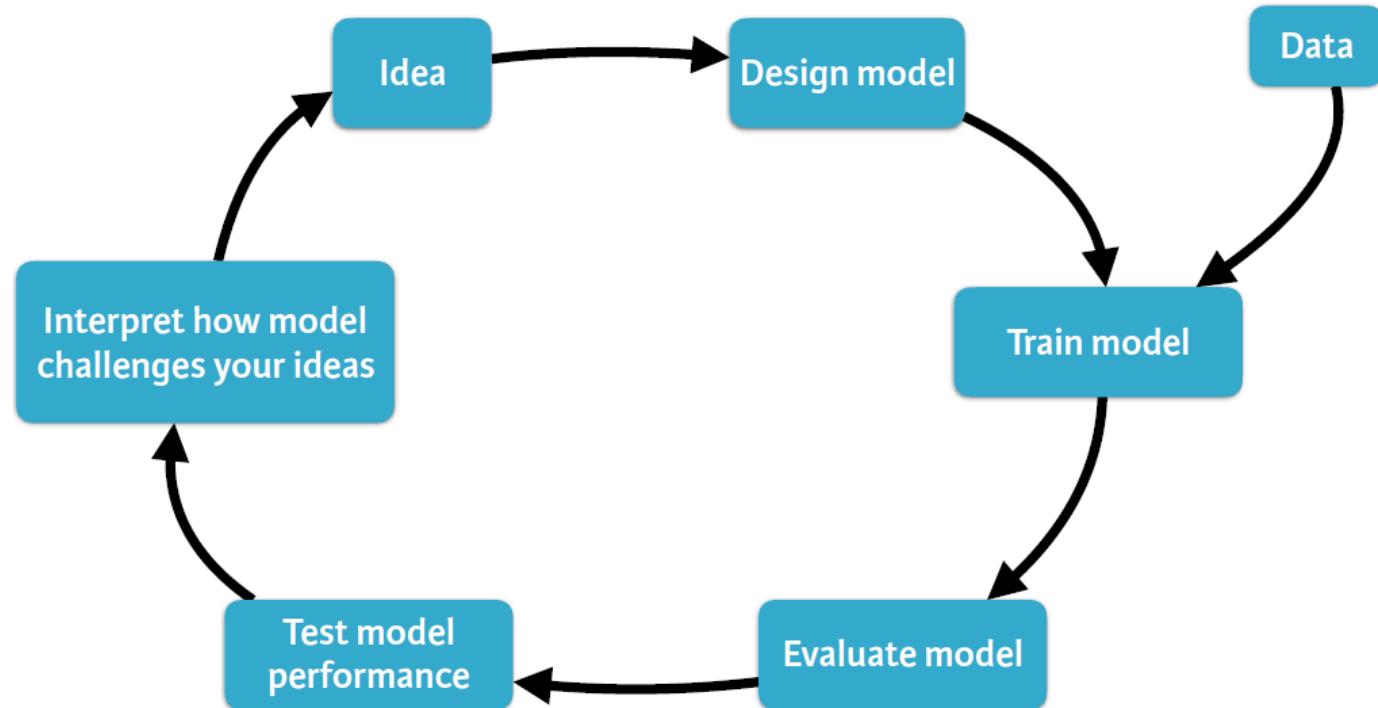
Question 05

- Which of the following is **not** an Auto ML tool?
 - A. Auto Keras
 - B. H2O Driverless AI
 - C. Scikit-learn
 - D. Amazon Lex

Data Science Pipeline



Modeling is a process



Pause and Reflect

- What is the process of Data Science? What can you do with it?
- Take a minute to think about that and move your eyes away from the screen.

Break for exercise...



Reality vs Expectation

- How is the Model
- Is it a good fit?
- Do we need to try another model?
- What variables are impacting the Output?

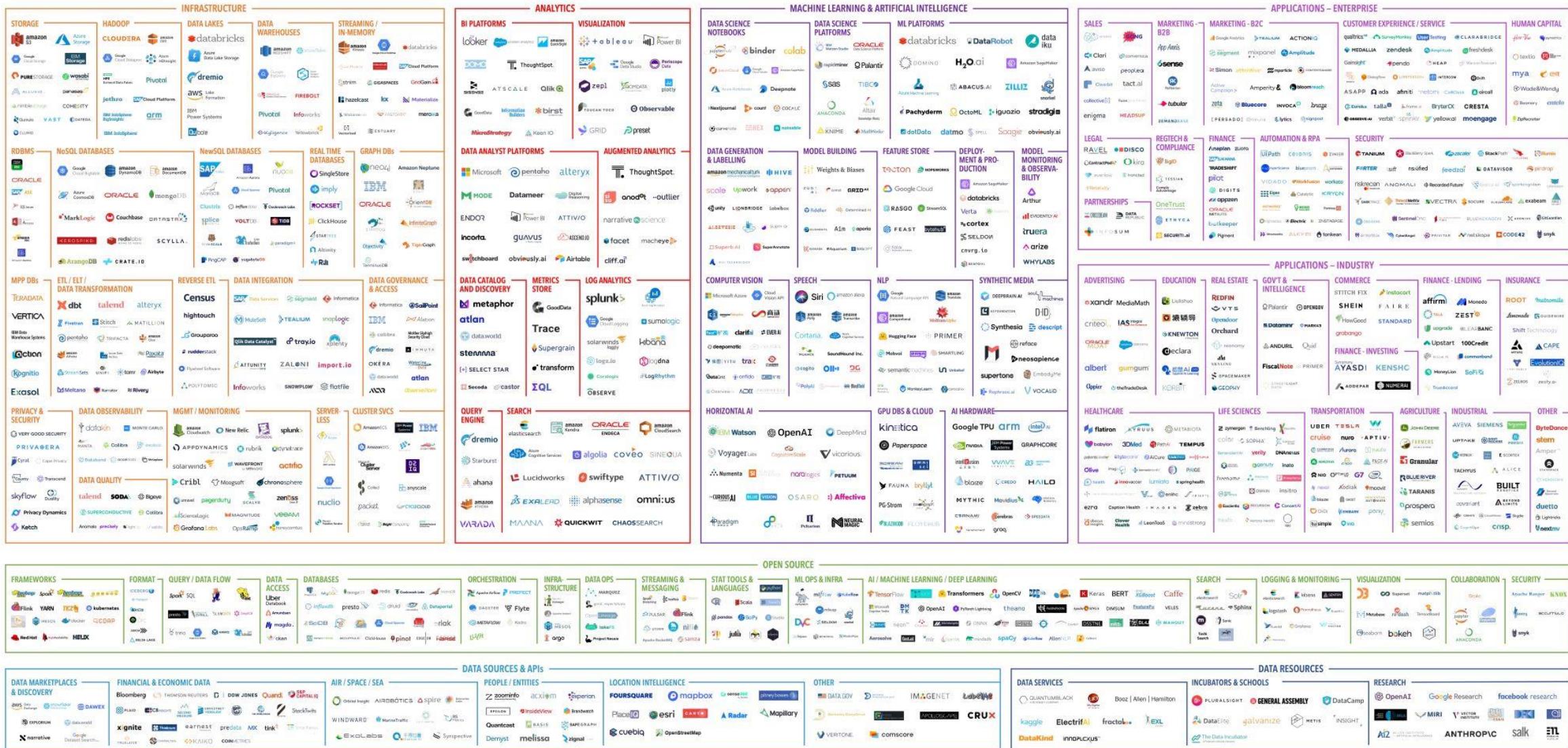
Which variables to consider?
What is the Data saying?

This is the analysis Data Scientist does **BUT** customer
Doesn't understand R-square, P-value etc..





MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, AND DATA (MAD) LANDSCAPE 2021



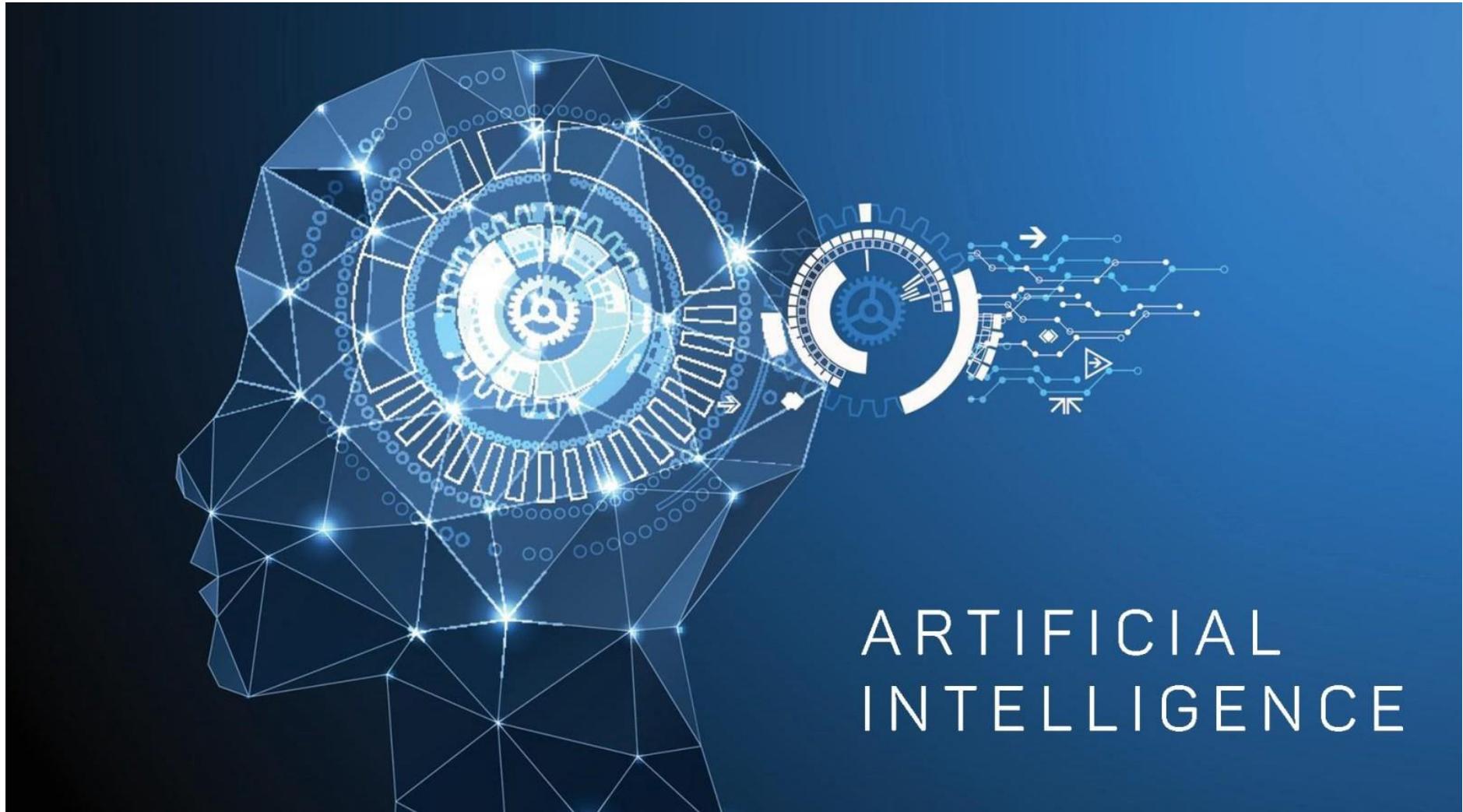
Version 3.0 - November 2021

© Matt Turck (@mattturck), John Wu (@john_d_wu) & FirstMark (@firstmarkcap)

mattturck.com/data2021

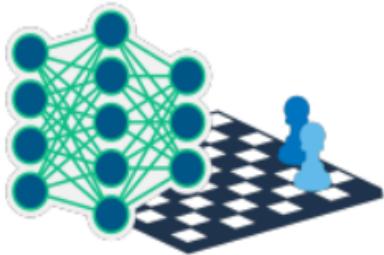
FIRSTMARK
EARLY STAGE VENTURE CAPITAL





ARTIFICIAL INTELLIGENCE

AI History



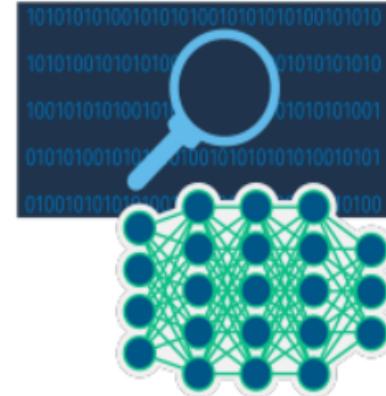
1950s-1970s Neural Networks

Early work with neural networks stirs excitement for “thinking machines.”



1980s-2010s Machine Learning

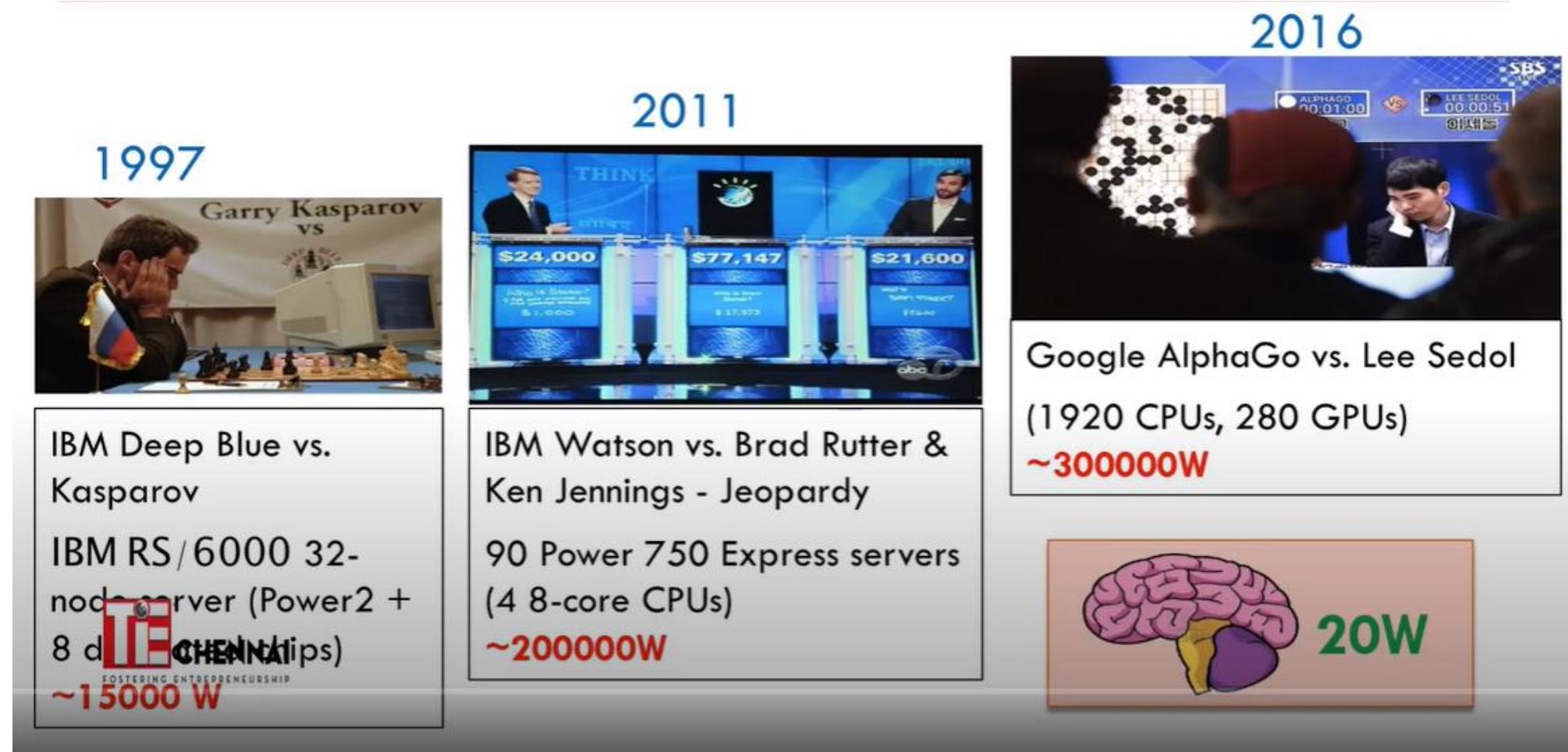
Machine learning becomes popular.



Present Day Deep Learning

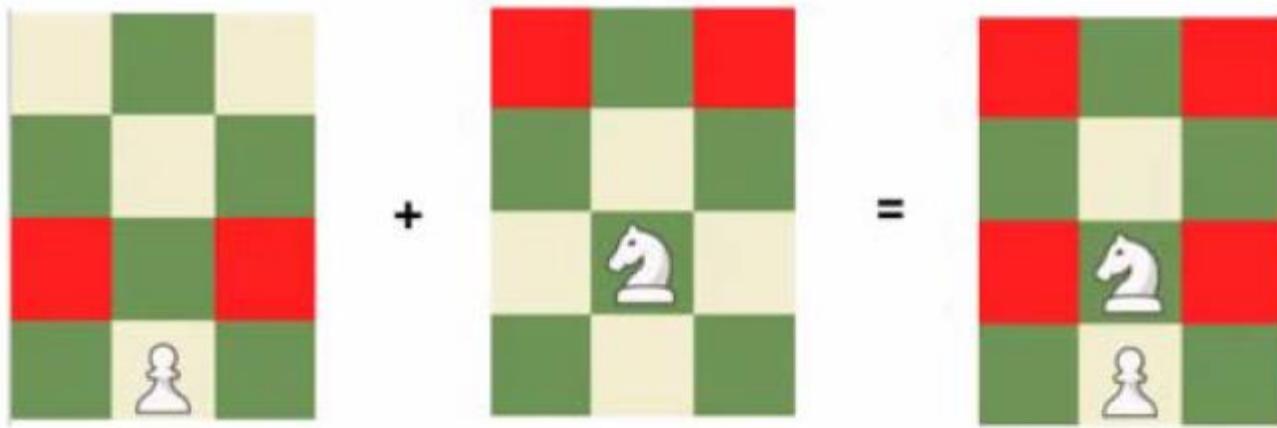
Deep learning breakthroughs drive AI boom.





Source: Artificial Intelligence- The Way ahead by Prof. V Kamakoti

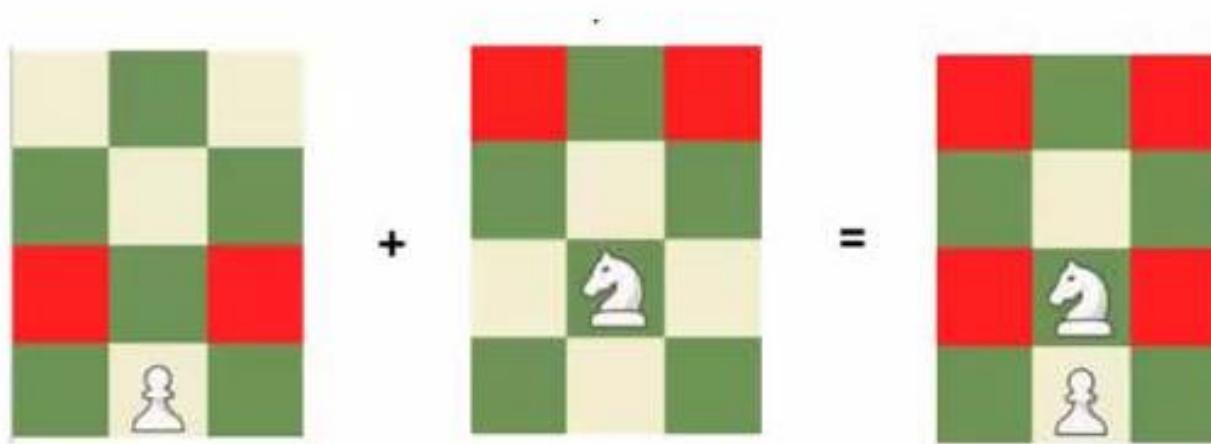
Ability to apply



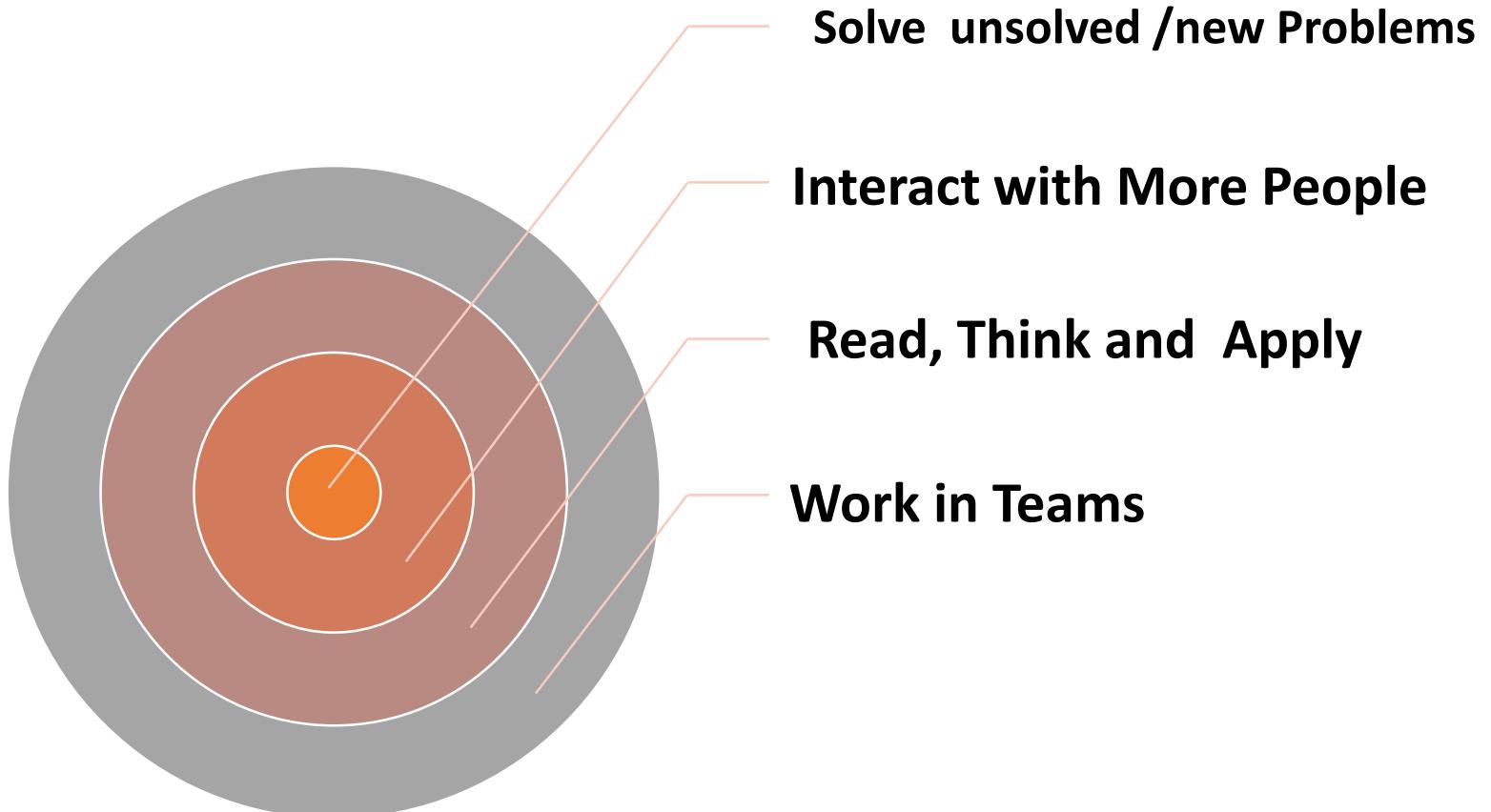
Ability to apply



Ability to apply



How to improve your analytical abilities

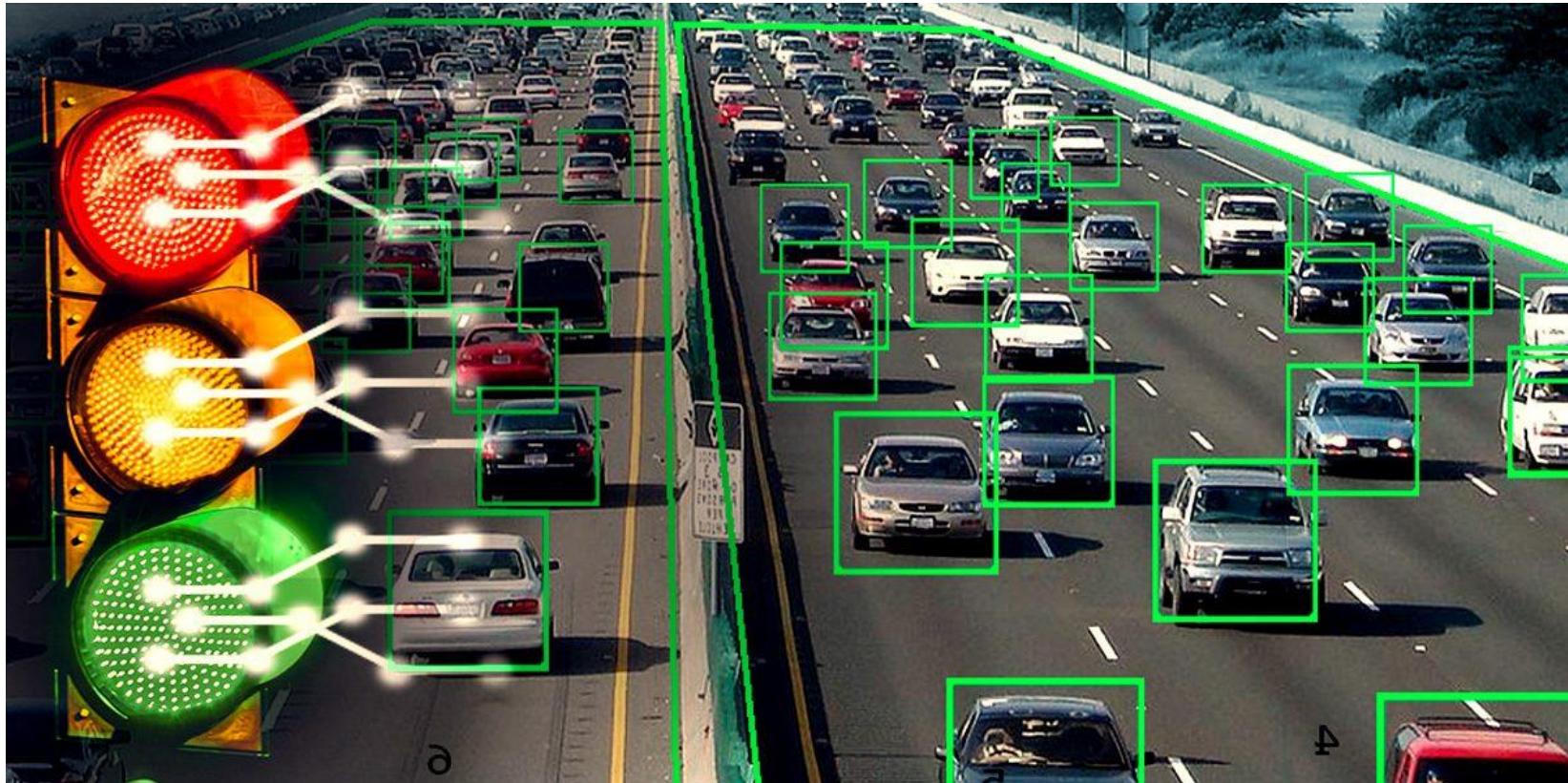


AI Developmental Units

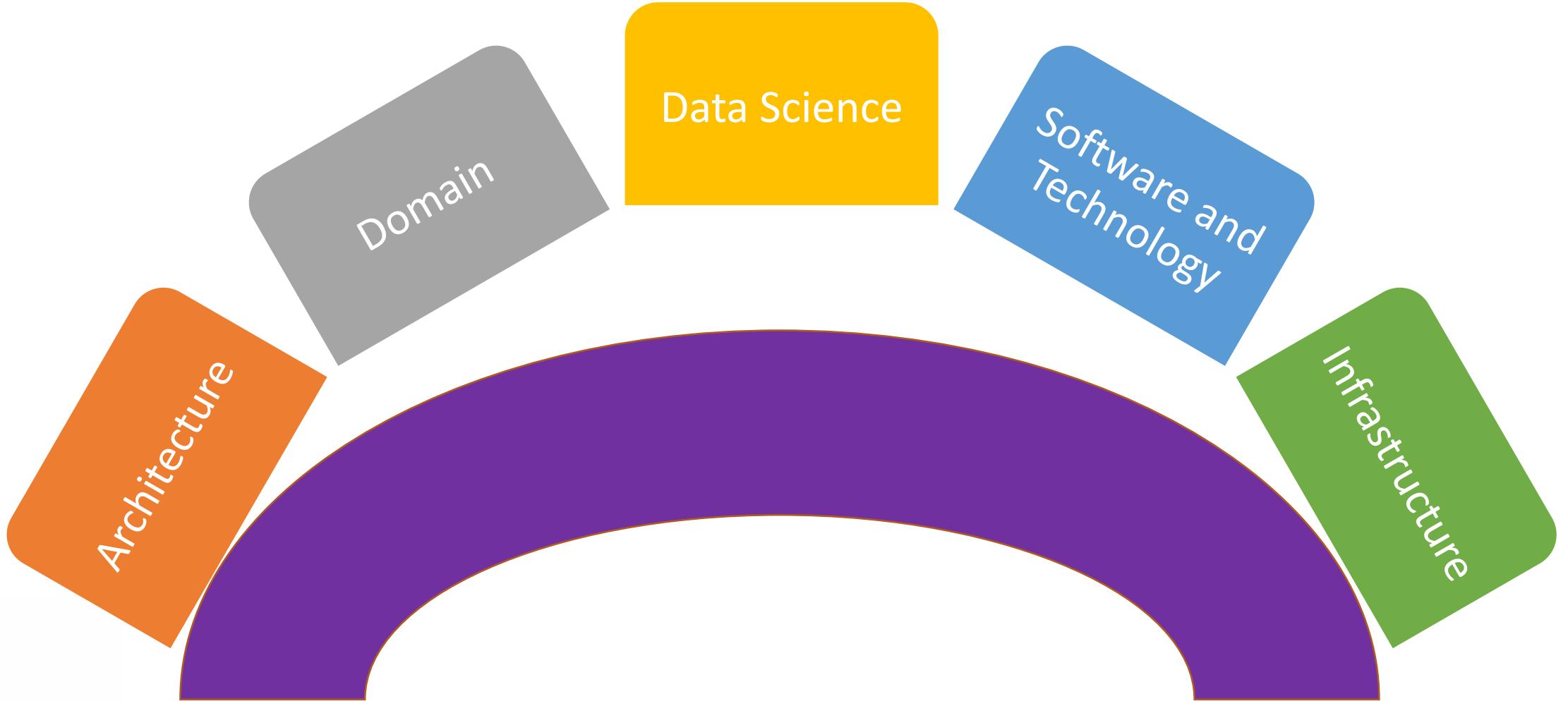
- Meet the wants
- Identify new use cases to meet customer DESIRE



Mobility sharing and Traffic Management



Challenge



Facebook AI over the years

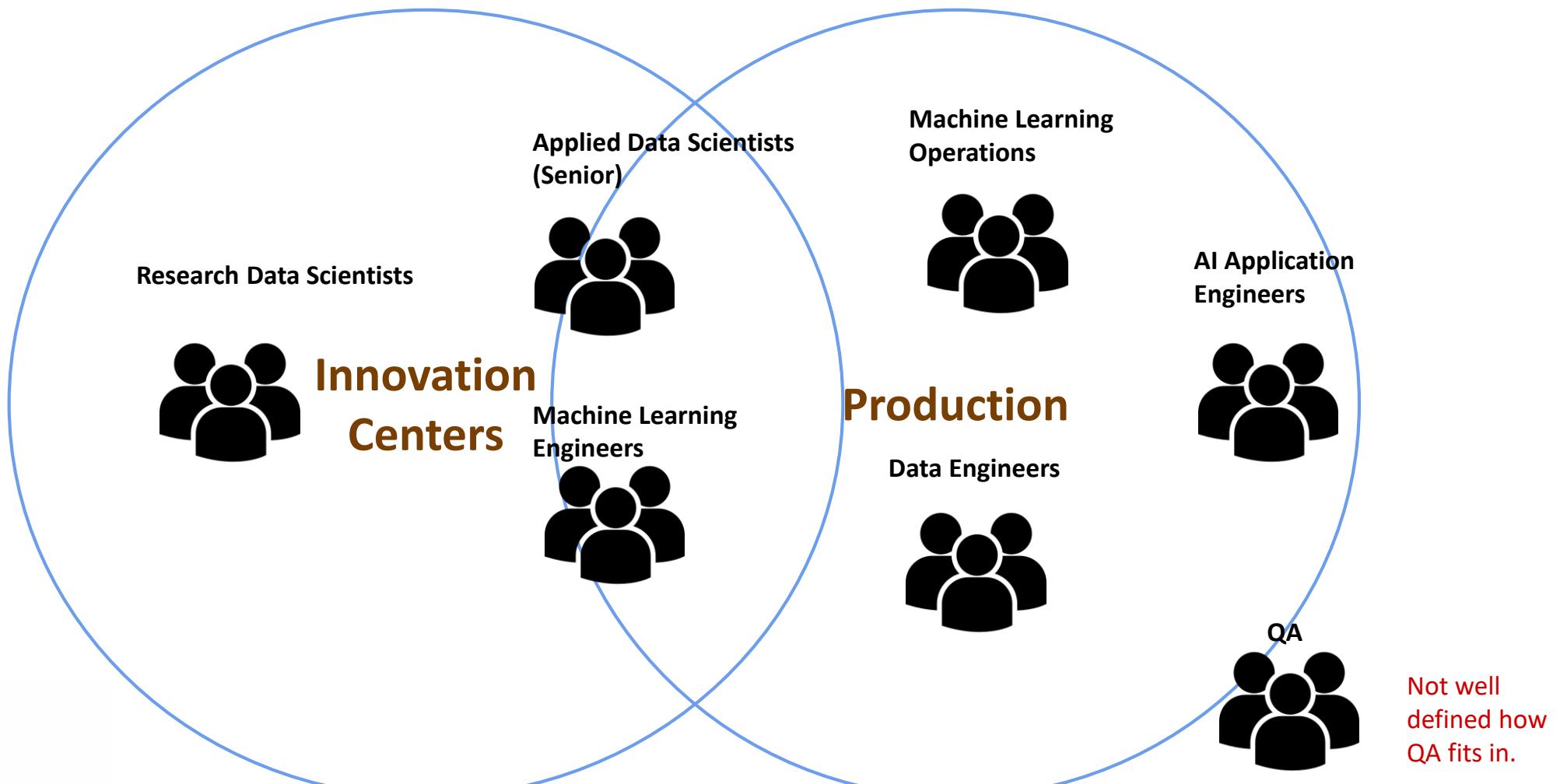


How to Manage scale?



Learn **Principles** and
Techniques

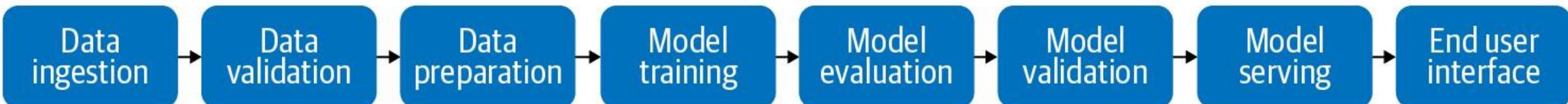
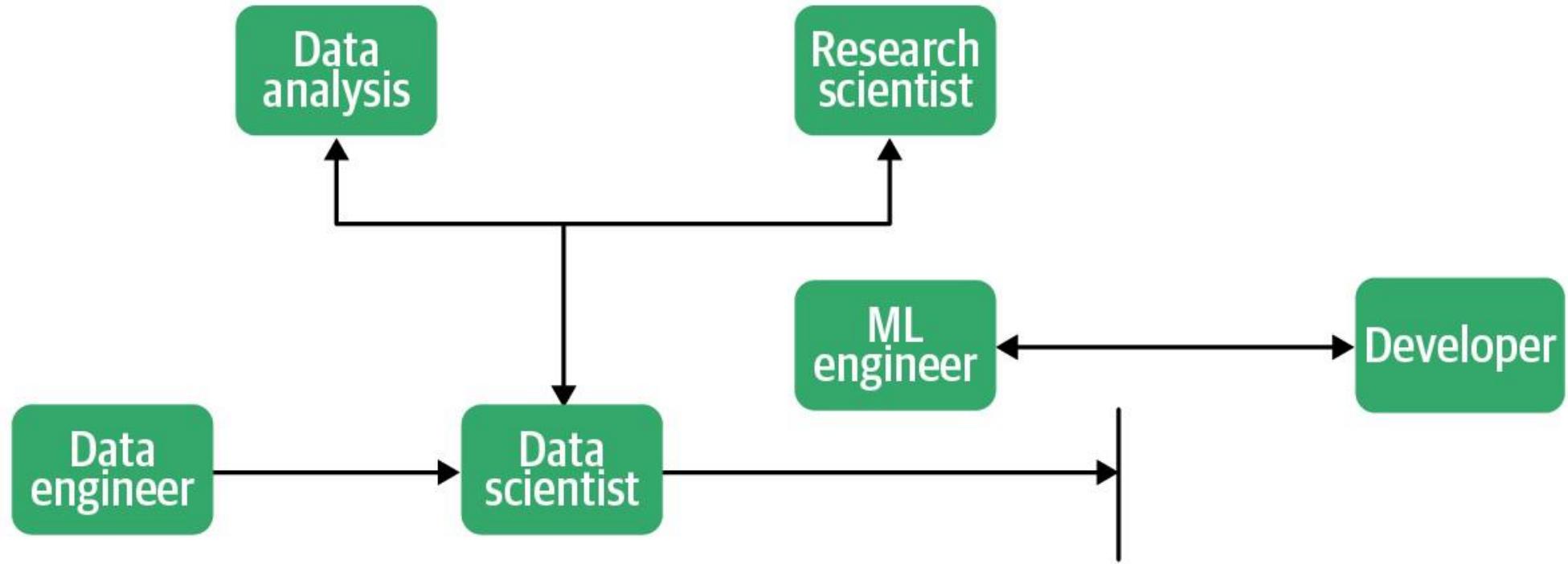
Job Definitions in ML and Data Science



Data Science Job roles

Business

Engineering



source: Machine Learning Design Patterns

Common Challenges in Data Science



Data Quality

Feature

- Measurement Error
- Recording error
- Duplicates
- Missing Features
- Label
 - Accuracy
 - Consistency
 - Data Set Completeness
 - Timeliness



PURPOSE

Reproducibility

Randomness
Framework
Dependencies

Drift Scale Objectives

DSRG @ IIT Kottayam

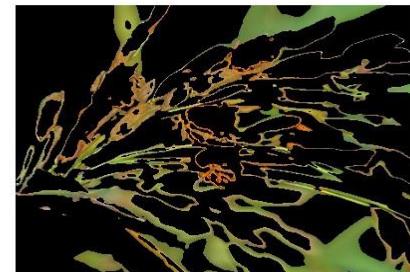
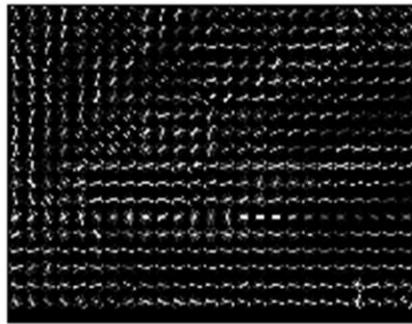
Data Science Research Group

Detection and Classification of Diseases in Rice crops

Input image



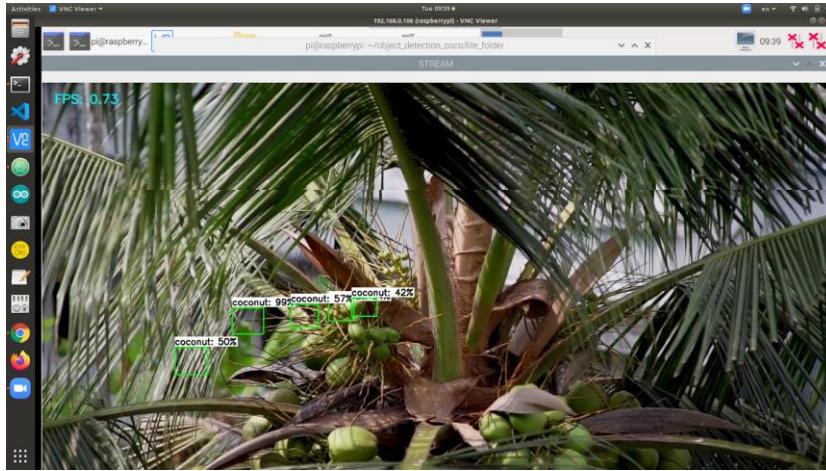
Histogram of Oriented Gradients



Left - Objects in cluster 1 and Right - Objects in cluster 2



Coconut Ripeness Detection using Computer Vision



Orzico
connecting people through high
speed opto-electric communication



Forest Inventory and Tree Measurement via 360° Spherical Images





Artificial Intelligence for Social Assessment (AIFORA)

