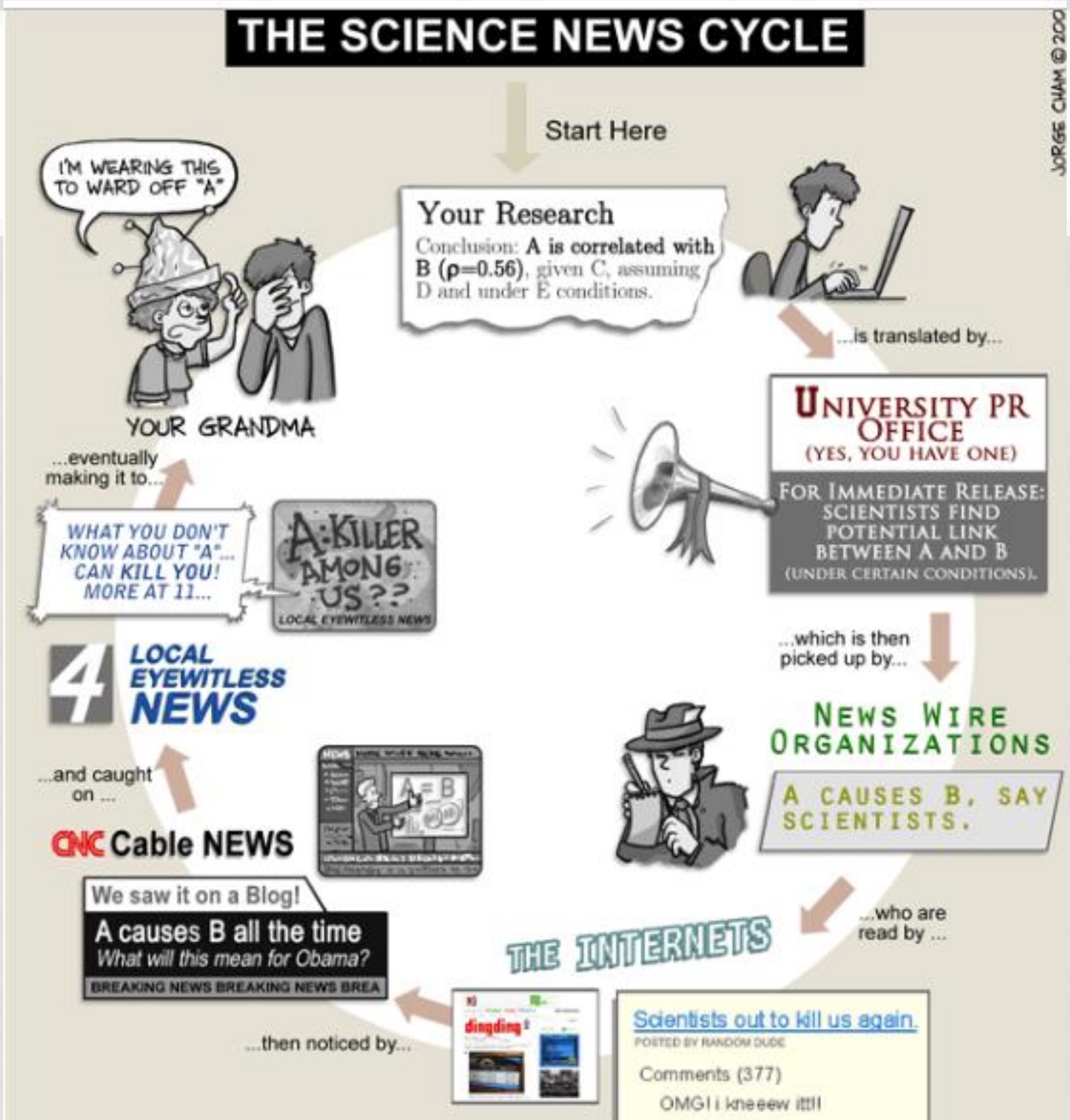
ARTIFICIAL INTELLIGENCE AND BIG DATA - 101


THE SCIENCE NEWS CYCLE
JORGE CHAM © 200
Start Here
Your Research
Conclusion: A is correlated with B (ρ=0.56), given C, assuming D and under E conditions.
...is translated by...
UNIVERSITY PR OFFICE
(YES, YOU HAVE ONE)
FOR IMMEDIATE RELEASE: SCIENTISTS FIND POTENTIAL LINK BETWEEN A AND B (UNDER CERTAIN CONDITIONS).
...which is then picked up by...
NEWS WIRE ORGANIZATIONS
A CAUSES B, SAY SCIENTISTS.
...who are read by ...
THE INTERNETS
Scientists out to kill us again.
POSTED BY RANDOM DUDE
Comments (377)
OMG! i kneeew it!!!
...then noticed by...
We saw it on a Blog!
A causes B all the time
What will this mean for Obama?
BREAKING NEWS BREAKING NEWS BREA
CNC Cable NEWS
...and caught on ...
4 LOCAL EYEWITLESS NEWS
WHAT YOU DON'T KNOW ABOUT "A"... CAN KILL YOU! MORE AT 11...
A: KILLER AMONG US??
LOCAL EYEWITLESS NEWS
...eventually making it to...
YOUR GRANDMA
I'M WEARING THIS TO WARD OFF "A"

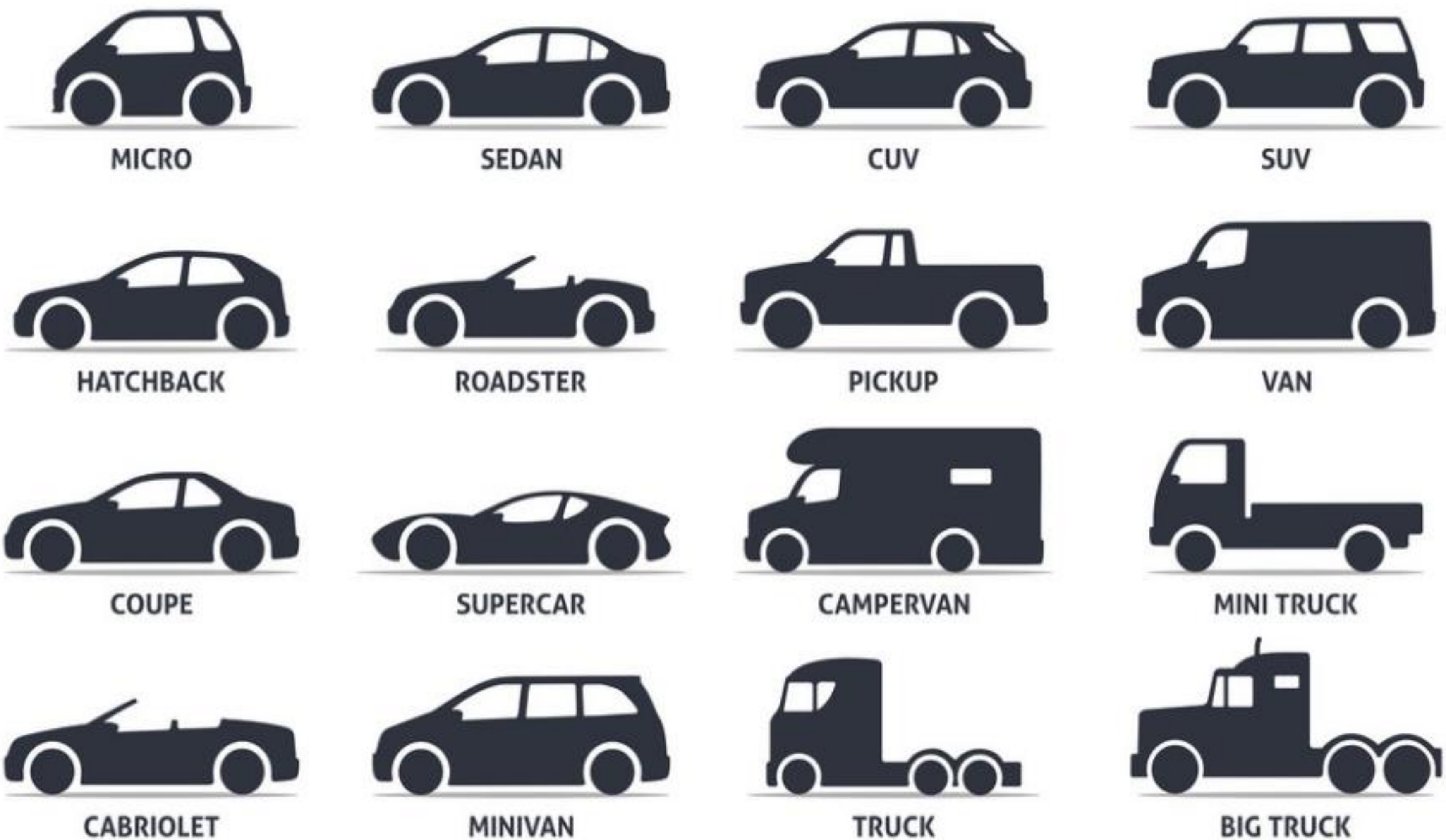
HONDA
Ford
mazda
HOLDEN
NISSAN
MITSUBISHI

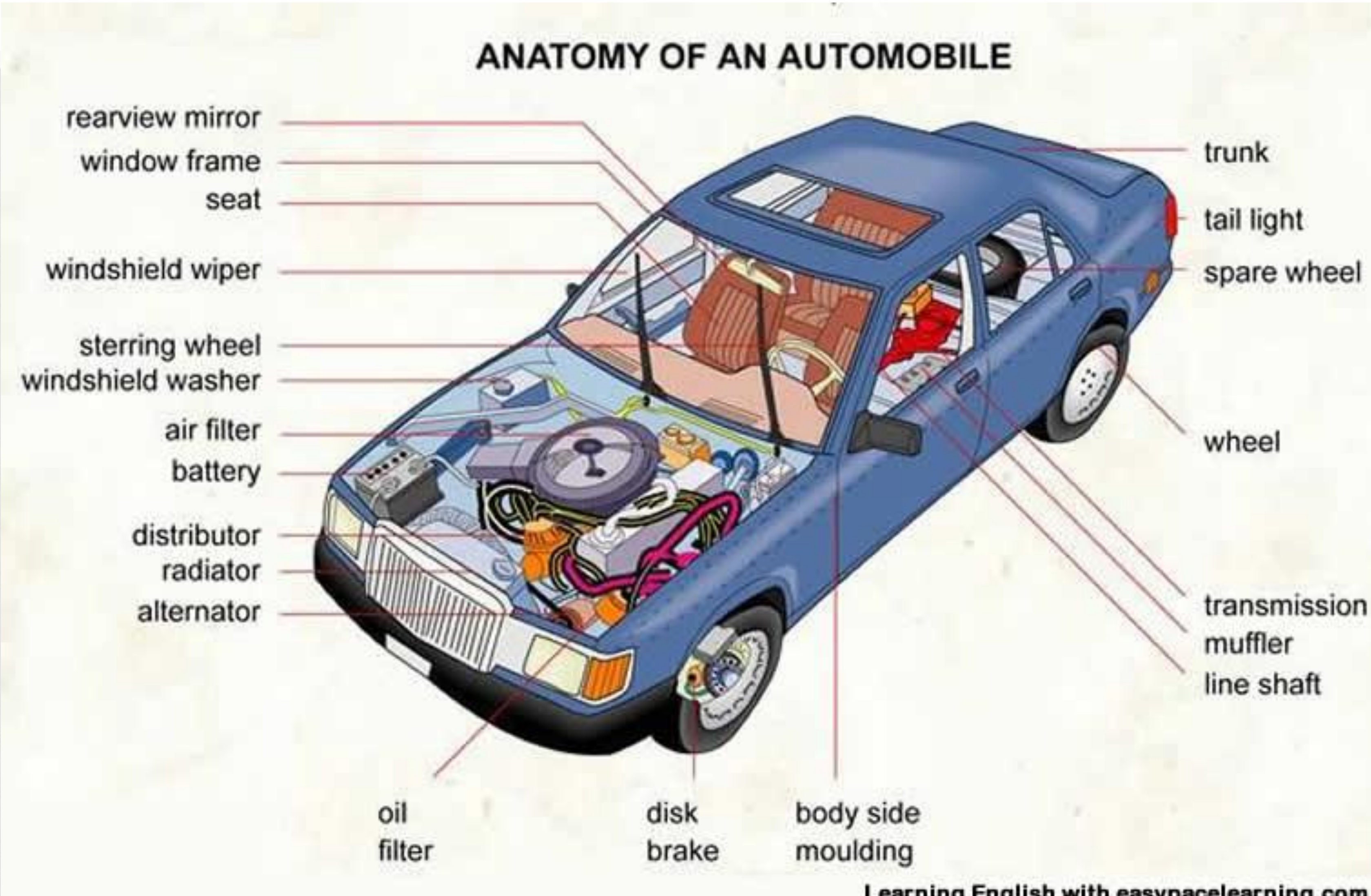
BMW
Audi
Jeep
DAEWOO
CHRYSLER
LEXUS

VOLVO
SUBARU
smart
SUZUKI
MINI
FIAT
LAND ROVER

CITROËN
SAAB
KIA
KIA MOTORS
The Power to Surprise
BENTLEY
MASERATI
ROVER

SSANGYONG
PORSCHE
STUTTGART
DAIHATSU
JAGUAR
ASTON MARTIN
ASTON MARTIN
ROLLS
ROYCE
Ferrari
ALFA ROMEO


MICRO
SEDAN
CUV
SUV
HATCHBACK
ROADSTER
PICKUP
VAN
COUPE
SUPERCAR
CAMPERVAN
MINI TRUCK
CABRIOLET
MINIVAN
TRUCK
BIG TRUCK



ANATOMY OF AN AUTOMOBILE
rearview mirror
window frame
seat
windshield wiper
sterring wheel
windshield washer
air filter
battery
distributor
radiator
alternator
oil filter
disk brake
body side moulding
trunk
tail light
spare wheel
wheel
transmission
muffler
line shaft
Learning English with easypacelearning.com

WHAT GOES INTO THE TYPE OF CAR YOU WANT?

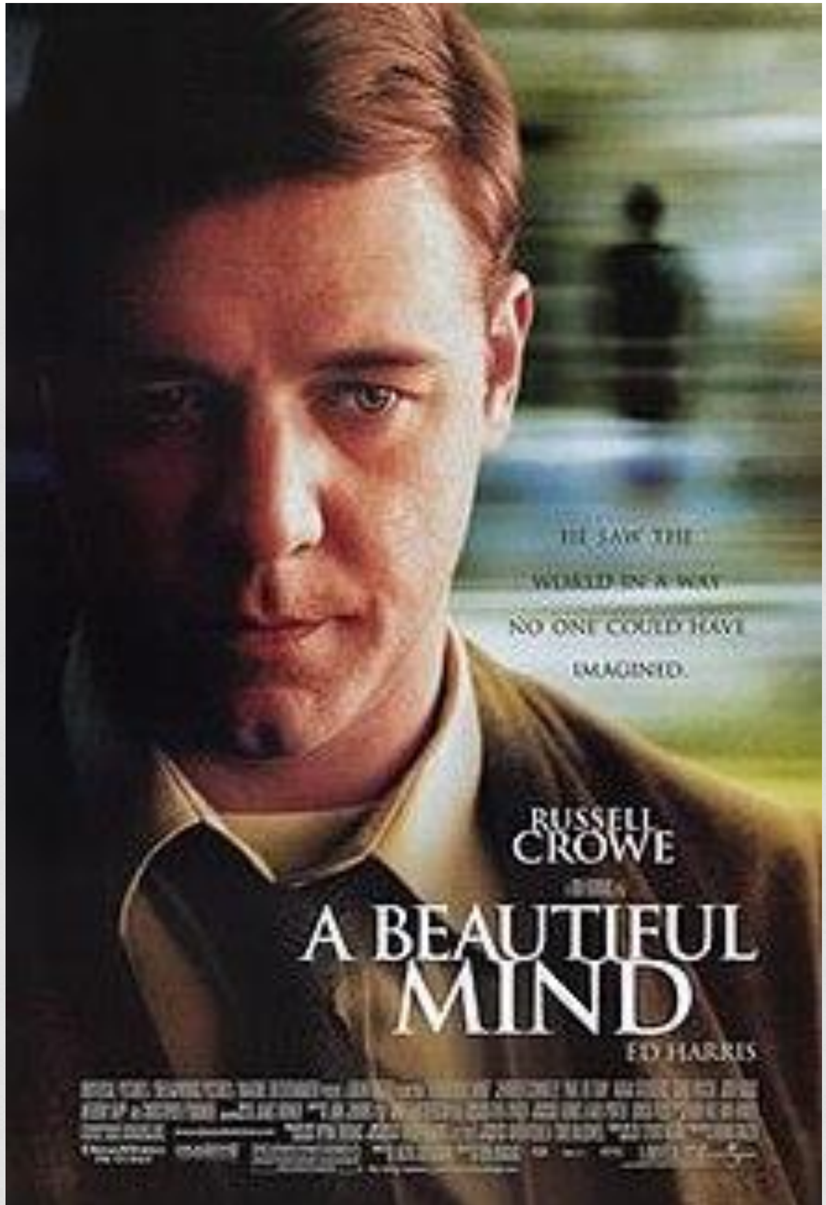

Content
History
Experience
Creative
STORYTELLING
Marketing
Share
Emotion
Communication

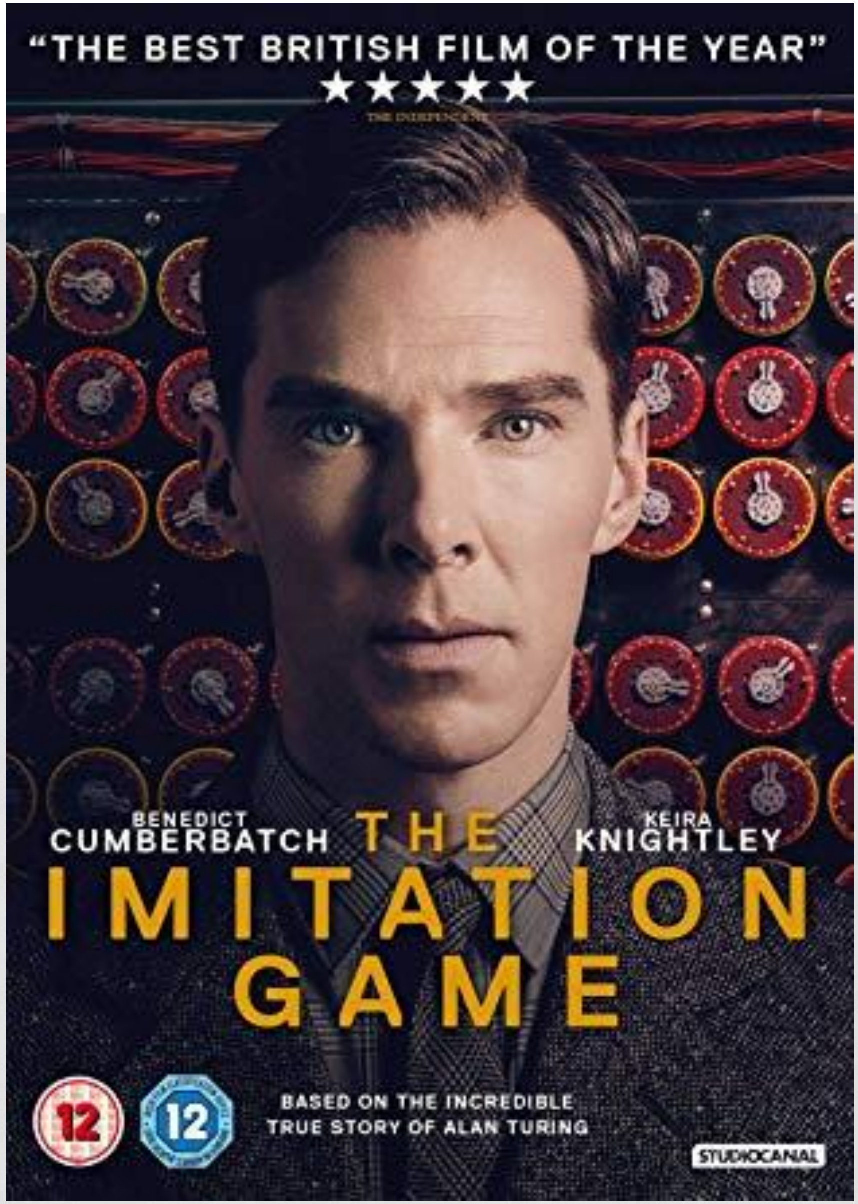
BASIC FORMULA TO GOOD STORYTELLING

BASIC FORMULA TO GOOD STORYTELLING

Breaking
Bad

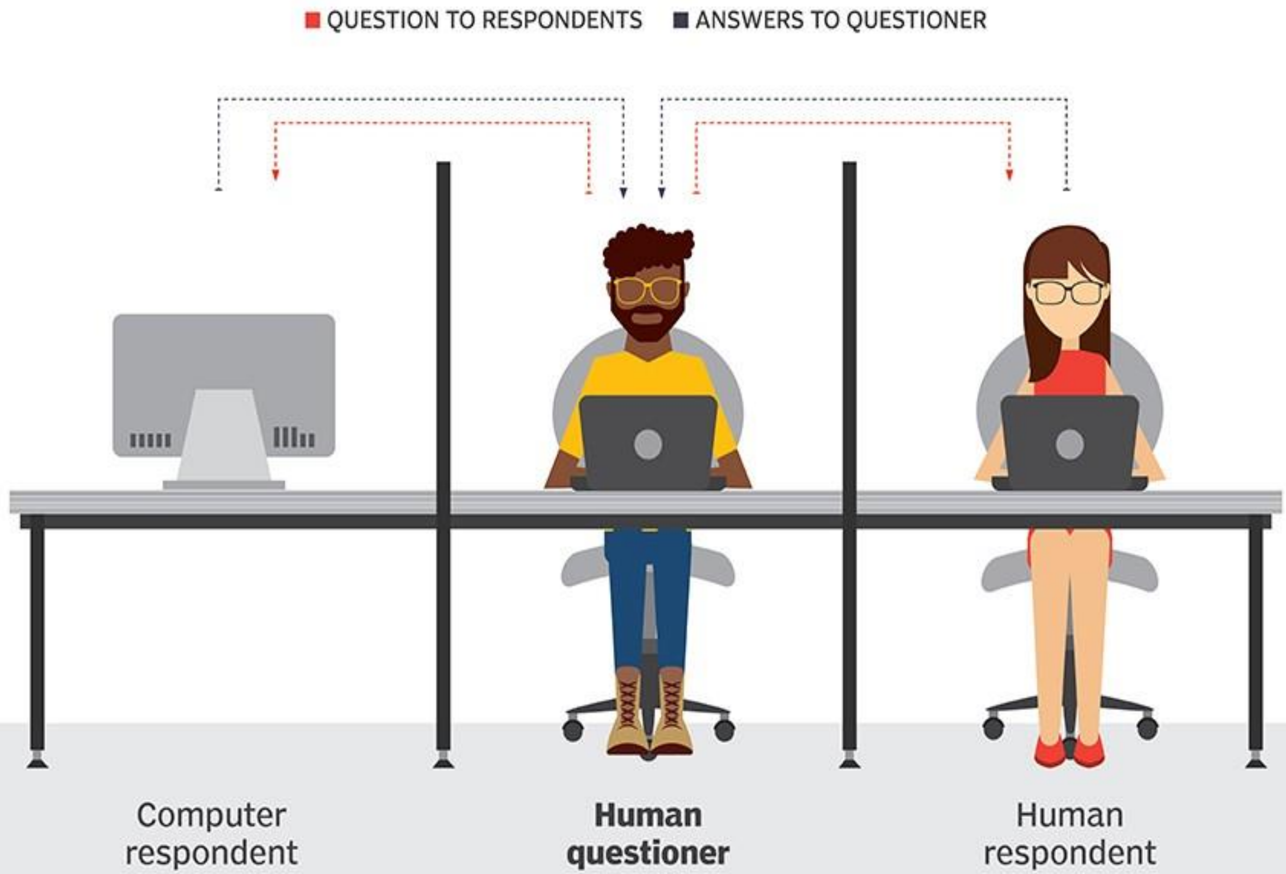

ROBERT DOWNEY JR. CHRIS HEMSWORTH MARK RUFFALO CHRIS EVANS SCARLETT JOHANSSON
BENEDICT CUMBERBATCH DON CHEADLE TOM HOLLAND CHADWICK BOSEMAN PAUL BETTANY
ELIZABETH OLSEN ANTHONY MACKIE SEBASTIAN STAN DANAI GURIRA LETITIA WRIGHT
DAVE BAUTISTA ZOE SALDANA JOSH BROLIN CHRIS PRATT
MARVEL STUDIOS
AVENGERS
INFINITY WAR
APRIL 27
IN REAL D 3D AND IMAX


HE SAW THE
WORLD IN A WAY
NO ONE COULD HAVE
IMAGINED.
RUSSELL
CROWE
A BEAUTIFUL
MIND
ED HARRIS


"THE BEST BRITISH FILM OF THE YEAR"
★★★★★
BENEDICT
CUMBERBATCH
KEIRA
KNIGHTLEY
THE
IMITATION
GAME
12
12
BASED ON THE INCREDIBLE
TRUE STORY OF ALAN TURING
STUDIOCANAL

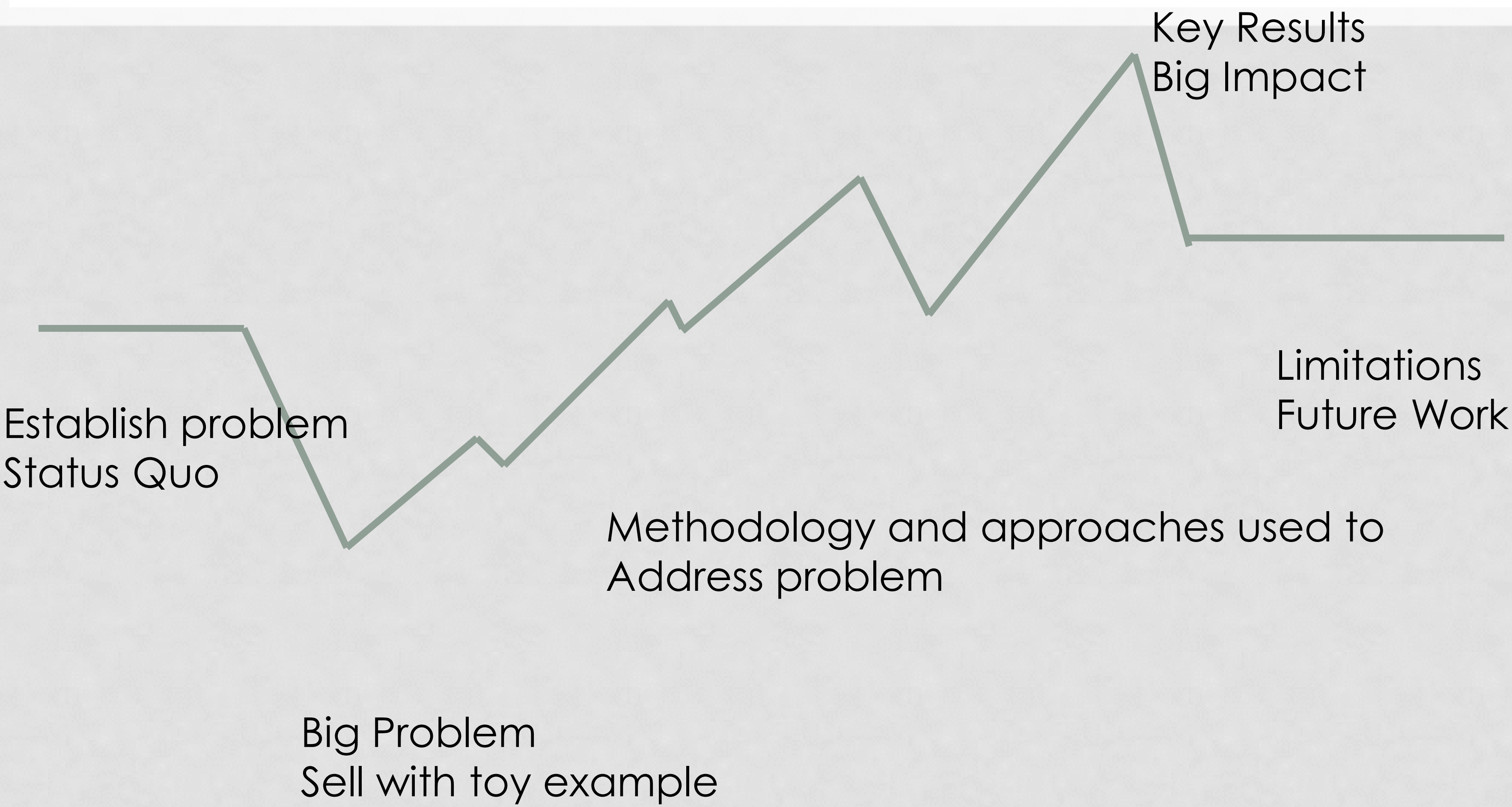
Turing test

During the Turing test, the human questioner asks a series of questions to both respondents. After the specified time, the questioner tries to decide which terminal is operated by the human respondent and which terminal is operated by the computer.




AN HBO ORIGINAL SERIES
GAME OF THRONES
THE COMPLETE FIRST SEASON
YOU WIN OR YOU DIE

BASIC FORMULA TO GOOD STORYTELLING WITH DATA



BIG DATA VS ARTIFICIAL INTELLIGENCE

- **Big Data** – Large Amounts of Data – **All About Efficient Computation**
  - Computer Architecture
  - Networking
  - Distributed Systems
  - Virtual Machines
  - Operating Systems
  - Parallel Computing
- **Artificial Intelligence** – Intelligent processing of data **– All About Smart Applied Math**
  - Statistics
  - Optimization
  - Probability Reasoning
  - Logic Reasoning
  - Machine Learning

BIG DATA

- How big is big data?
  - If you can't process the data quickly on your PC, it is big data
- Why is that the benchmark?
  - Because you have to start being mindful of how to process and manage data.
  - Opens up another can of worms

HOW DO YOU PROCESS DATA FASTER?

- How to cut a lawn faster?
- How to cook more food?
- How to get to work faster?

Add more resources
Add better resources
Parallelization

BUILDING BLOCKS OF COMPUTATION

- **Computation Power ($$$$)**
  - CPU (Central Processing Unit) – cores
    - More cores, more ability to parallel process
    - More CPUs more cores
    - Able to do general computation
  - GPU (Graphic Processing Unit) – cores
    - Has many more cores than a typical CPU (100 vs 8)
    - Only does specific computations – matrix, vector, computation
- **Memory**
  - RAM ($$) - Orders of Magnitude faster than reading from disk
  - Hard Disk (pennies) – Faster than hitting network for more resources
  - Solid state drive (tens of pennies) – Faster than hard drives, slower than RAM
- **Network ($$ - need people to manage) – distributed computing**
  - Ability to access more resources (SLOW, but relatively cheap money wise)
    - Computers with RAM, ROM, and CPU
    - Algorithms must be in place in order for these different computers to communicate

PARALLEL COMPUTATION (COMPUTER ARCHITECTURE AND OPERATING SYSTEM)

- Multi-t
- Multi-p
- Fork pr
- Semap
- Locking
- Mutexe
- CPU Pi

Computer Scientist, Programmers, and Engineers SUCK at Parallel Computing

Big Data Technologies Make parallel Computation Easier for Everyone

https://en.wikipedia.org/wiki/Computer_architecture

https://en.wikipedia.org/wiki/Operating_system

GOOGLE STARTED MODERN BIG DATA MOVEMENT (FINANCIAL DECISION)

- Database companies would sell huge machines to process large amounts of data (Supercomputers)
  - CPU with 100 cores
  - 100GB of RAM
  - 10TB Hard drive
  - VERY EXPENSIVE

HOW DID BIG DATA GET STARTED

Lenovo 6241HDU System x3950 X6 Server

Manufacturer: Lenovo **UPC:** 889488019632 **SKU:** 6241HDU **Condition:** New

✔ In Stock

★★★★★ Write Review

$192,240.23

1

ADD TO CART

CHEAPER TO USE MANY SMALLER COMPUTERS IN TANDEM THAN ONE LARGE COMPUTER

DECEMBER 2, 2010 WEBLOG

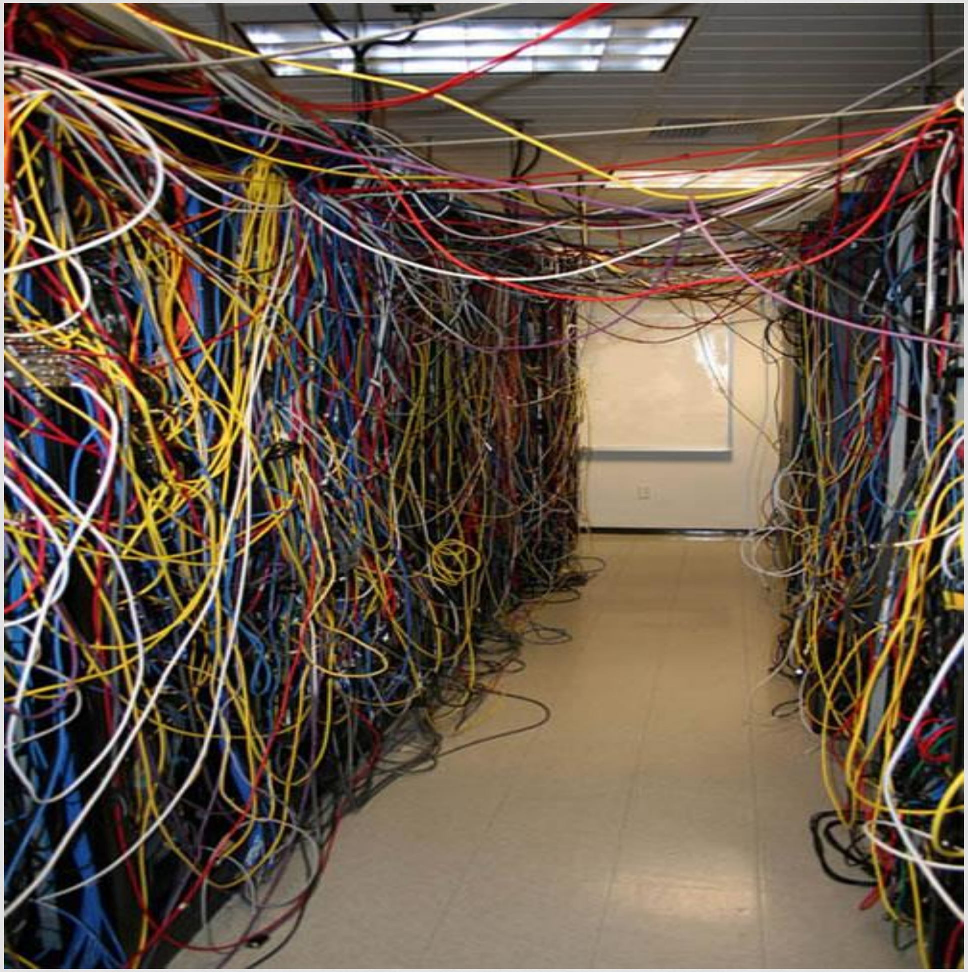
US Air Force connects 1,760 PlayStation 3's to build supercomputer

by Lisa Zyga , Phys.org

The Condor Cluster consists of 1,760 Sony PlayStation 3's, and is the US Department of Defense's fastest inte…

(Phys.org)—About the 33rd largest supercomputer in the world right now is the US Air Force Research Laboratory's (AFRL) newest system, which has a core made of 1,760 Sony PlayStation 3 (PS3) consoles. In addition to its large capacity, the so-called "Condor Cluster" is

PUTTING LOTS OF SMALL COMPUTERS TOGETHER TO DO FAST COMPUTATION CAUSES PROBLEMS

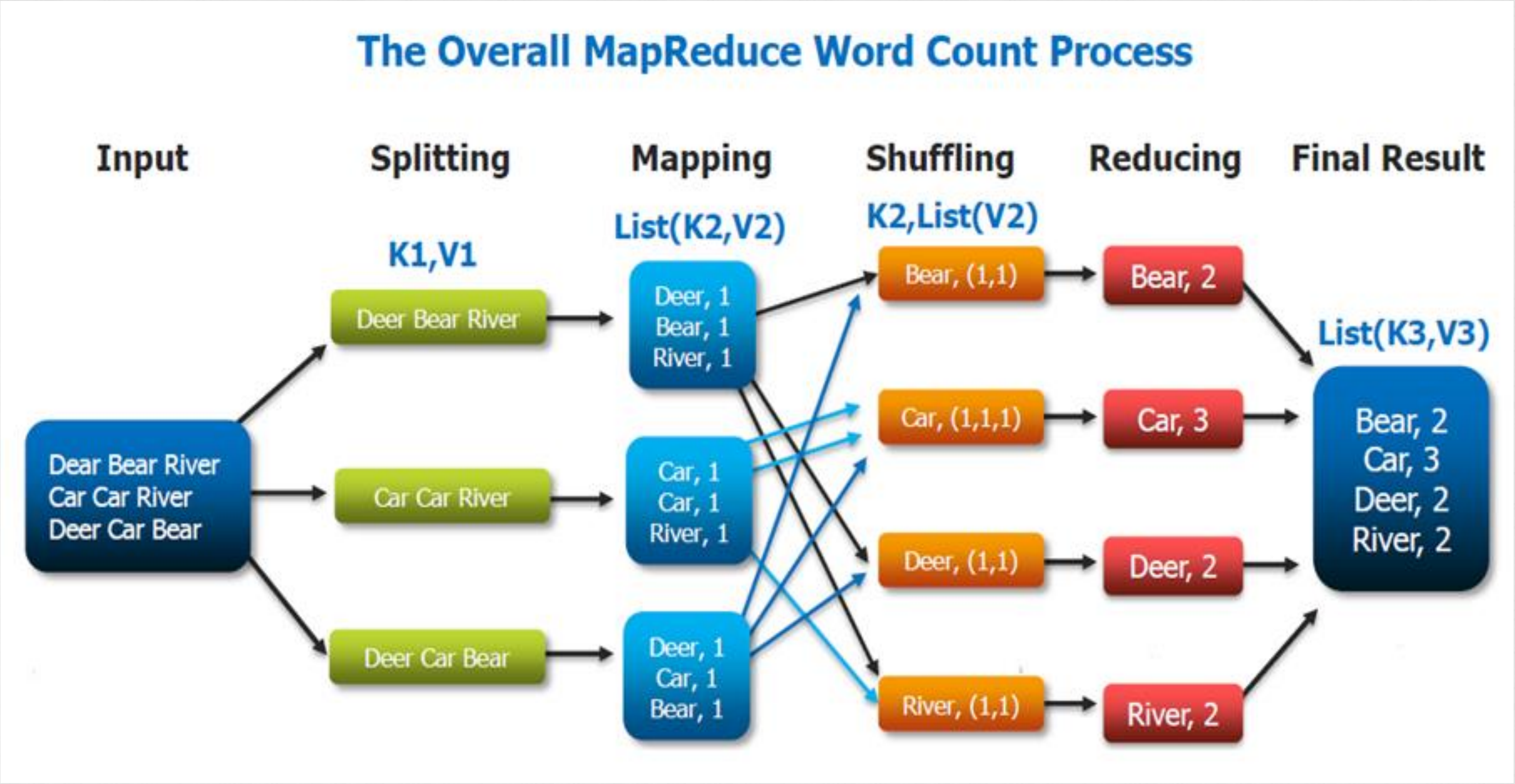


WHAT KIND OF PROBLEMS?

- Server failures – nodes not responding
- Connectivity – network failures
- Disk failures – loss of data
- Distribution of work/jobs – how to distribute data
- Security
- Failur … ilure
  - Ava
  - Faul
    - Ex
    - C
  - Auto Scaling (on Demand processing)

Cloud Service Providers (GCP, AWS, Azure) Abstract this Complexity Away

MAPREDUCE – DISTRIBUTED PROGRAMMING MODEL



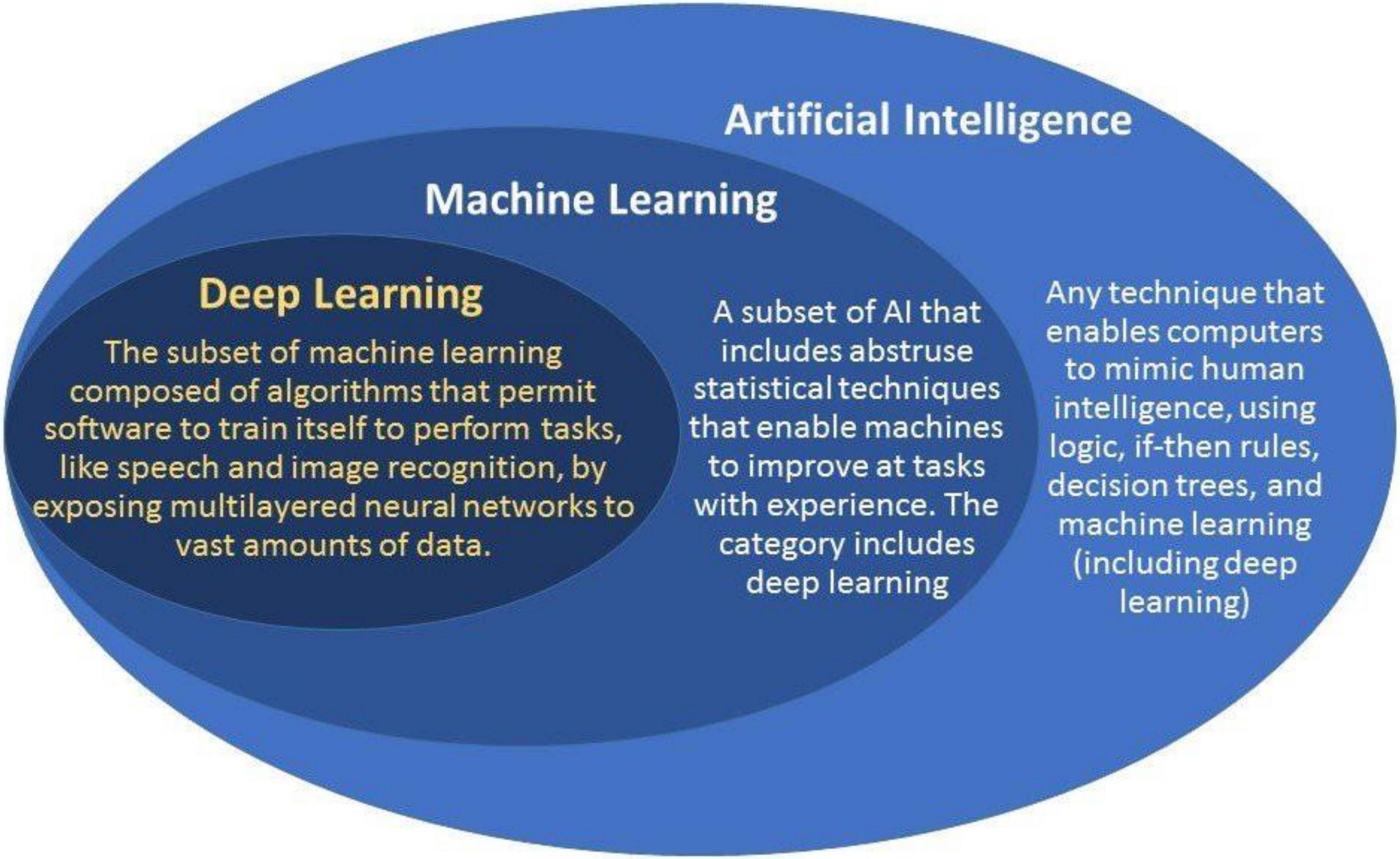
SPARK VS HADOOP

Parameters	**Spark**	**Hadoop**
Fault Tolerance	Spark RDD guarantees fault tolerance	Uses replication for fault tolerance
Speed	Faster due to in-memory computation	Relatively slower than Spark
OS Support	Linux, Windows, macOS	Linux, Windows, macOS
High Level Language	Python, Scala, Java, R, Spark SQL	Java, Python, Pig, Hive SQL
Machine Learning	Has its own set of ML libraries	Requires interfacing with other libraries

BIG DATA TECHNOLOGY

- Bound by Resources
  - How many servers?
  - How many CPUs?
  - How many cores per CPU?
  - How much RAM?
  - How much Disk?

ARTIFICIAL INTELLIGENCE



ARTIFICIAL INTELLIGENCE

- Logical Reasoning
  - If A-> B, B->C, then A->C
- Search
  - Shortest Path Problems
  - Constraint Satisfaction Problems
- Probabilistic Reasoning
  - Naïve Bayes Classification
  - MAP Decision
- Machine Learning
  - Unsupervised learning
  - Supervised Learning

UTILITY FUNCTION

- A **utility function** is able to represent those preferences if it is possible to assign a real number to each alternative, in such a way that *alternative a* is assigned a number greater than *alternative b* if, and only if, the individual prefers *alternative a* to *alternative b*.
- Mathematical function for determining the value of one option versus another.

UTILITY FUNCTIONS FOR DECISION MAKING

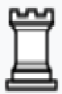
- Artificial intelligence and machine learning usually starts with defining a utility function that represents the desirability of an outcome.
- Determining the best utility function is subjective but you do need to define one that is "reasonable".
- Quantifying options is "human work".
  - Not always straightforward


MARVEL STUDIOS
AVENGERS
ENDGAME

CHESS

Standard valuations [edit]

The following table is the most common assignment of point values (Capablanca & de Firmian 2006:24–25), (Seirawan & Silman 1990:40), (Soltis 2004:6), (Silman 1998:340), (Polgar & Truong 2005:11).

Symbol					
Piece	pawn	knight	bishop	rook	queen
Value	1	3	3	5	9

Value of non-passed pawn in the opening

Rank	a & h file	b & g file	c & f file	d & e file
2	0.90	0.95	1.05	1.10
3	0.90	0.95	1.05	1.15
4	0.90	0.95	1.10	1.20
5	0.97	1.03	1.17	1.27
6	1.06	1.12	1.25	1.40

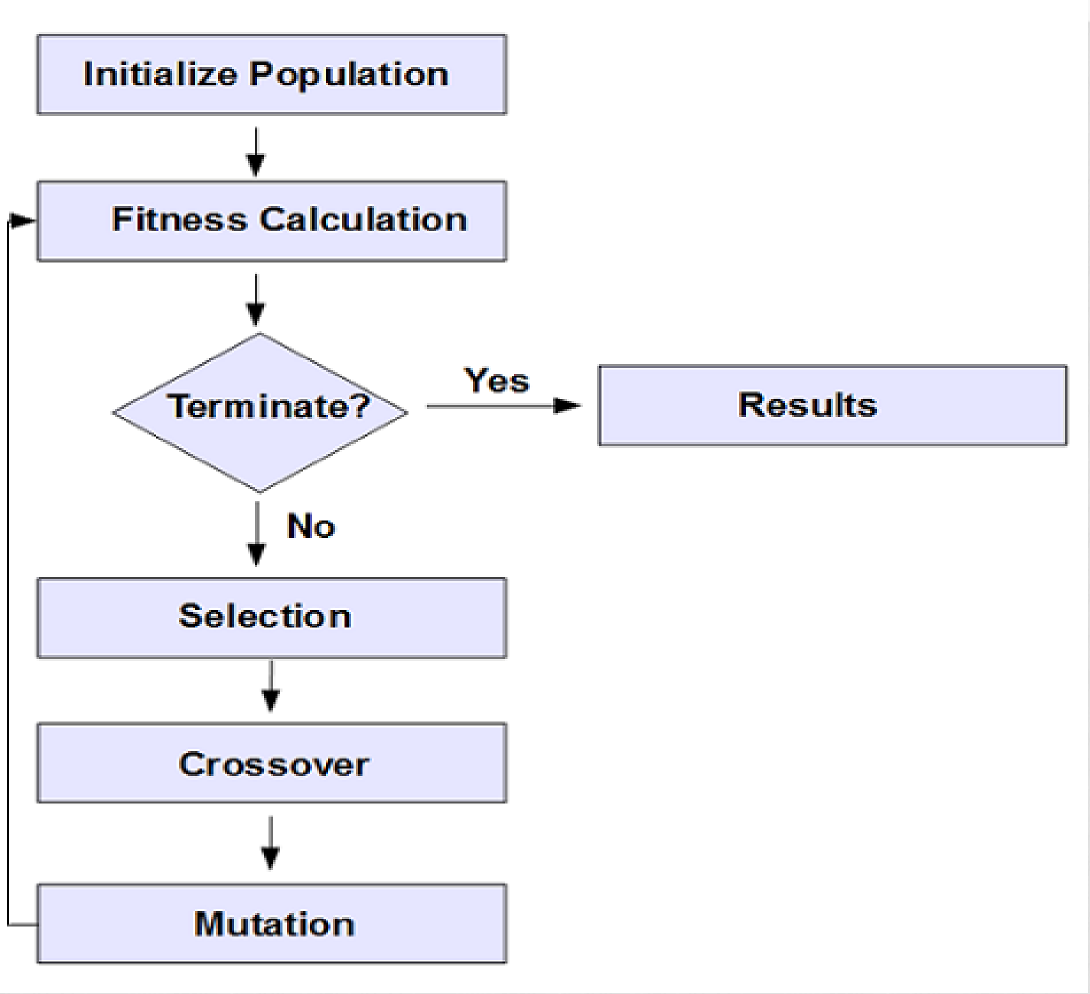
Value of non-passed pawn in the endgame

Rank	a & h file	b & g file	c & f file	d & e file
2	1.20	1.05	0.95	0.90
3	1.20	1.05	0.95	0.90
4	1.25	1.10	1.00	0.95
5	1.33	1.17	1.07	1.00
6	1.45	1.29	1.16	1.05

ARTIFICIAL INTELLIGENCE IN CHESS

- Given all possible outcomes for the next few moves – which move gives me the best chance of highest the highest value of my pieces (according to my utility function).
  - Solved Via Exhaustive search of All possibilities to see what works
  - Computers struggled for years because it couldn't represent the future well in memory (only next 2 moves vs next 1000 moves)
  - With supercomputers, these additional combinations can be stored and checked.

GENETIC ALGORITHMS



https://rednuht.org/genetic_cars_2/

OPTIMIZATION

- Trying to find best option (parameters) which maximizes or minimizes a utility function.
  - If you can't optimize intelligently, you do exhaustive search
    - **Exhaustive search** compares each possible option, computes the utility function, and select the option that maximizes/minimizes it (**Grid search**).
  - Why not use **exhaustive search**?
    - Usually too many options to check – even with supercomputers
  - Alternatives – **Optimization algorithms** (complex)

EXAMPLE

- How to determine a utility function for determining the best time to leave to go the airport?

EXAMPLE

- What Cuisine Should I Eat For Dinner Tonight?

- Table lookup – Burgers = 1.00
- Pizza – 1.00
- Mexican = 0.9
- Gyros = 0.6
- Yogurt = 0.2
- Argmax(Food) = Food_type - Food_type(recent) + rand(n)

LOGICAL REASONING – KNOWLEDGE BASE

A **knowledge-based system** (**KBS**) is a computer program that reasons and uses a knowledge base to solve complex problems. The term is broad and refers to many different kinds of systems. The one common theme that unites all knowledge based systems is an attempt to represent knowledge explicitly and a reasoning system that allows it to derive new knowledge. Thus, a knowledge-based system has two distinguishing features: a knowledge base and an inference engine.

Logical Reasoning - Validity and soundness

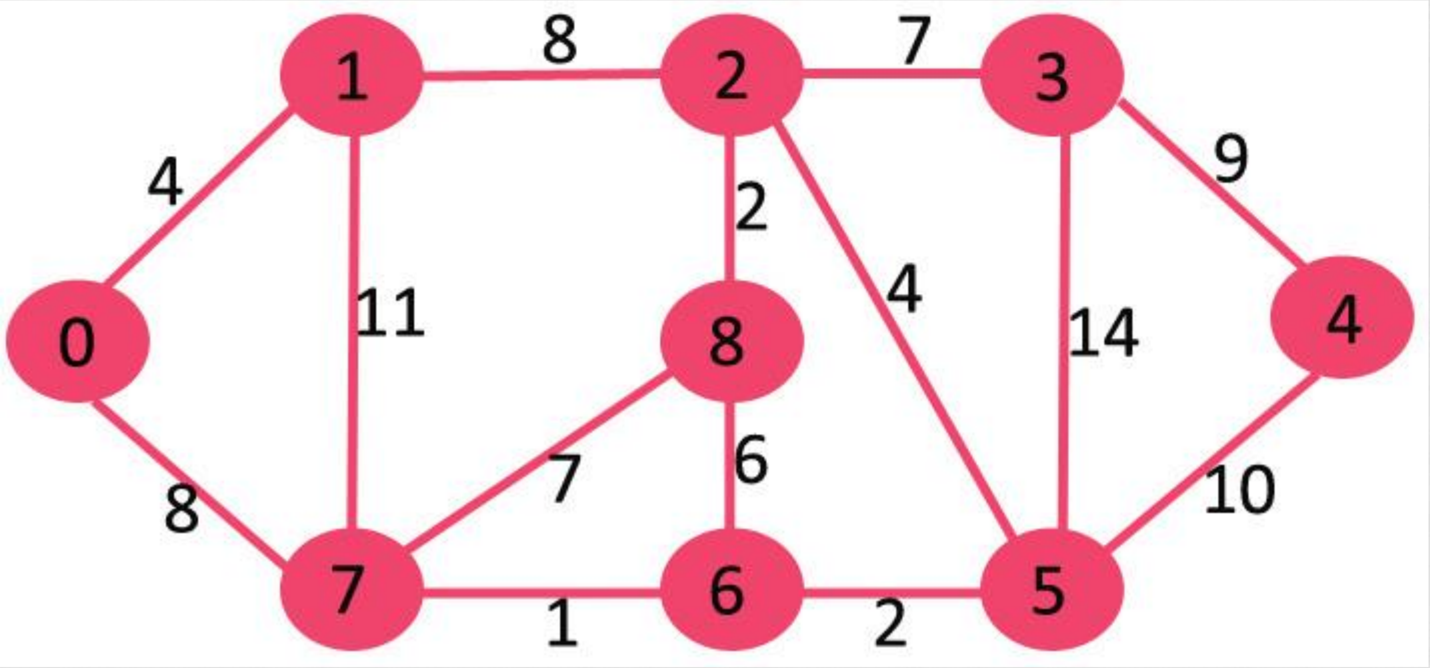
Argument terminology
Deductive arguments are evaluated in terms of their *validity* and *soundness*.
An argument is “**valid**” if it is impossible for its premises to be true while its conclusion is false. In other words, the conclusion must be true if the premises are true. An argument can be “valid” even if one or more of its premises are false.
An argument is “**sound**” if it is *valid* and the premises are true.
It is possible to have a deductive argument that is logically *valid* but is not *sound*. Fallacious arguments often take that form.
The following is an example of an argument that is “valid”, but not “sound”:
1. Everyone who eats carrots is a quarterback.
2. John eats carrots.
3. Therefore, John is a quarterback.

IBM WATSON

IBM WATSON

- Watson was created as a question answering (QA) computing system that IBM built to apply advanced natural language processing, information retrieval, knowledge representation, automated reasoning, and machine learning technologies to the field of open domain question answering.[2]
- The key difference between QA technology and document search is that document search takes a keyword query and returns a list of documents, ranked in order of relevance to the query (often based on popularity and page ranking), while QA technology takes a question expressed in natural language, seeks to understand it in much greater detail, and returns a precise answer to the question.[12]
- When created, IBM stated that, "more than 100 different techniques are used to analyze natural language, identify sources, find and generate hypotheses, find and score evidence, and merge and rank hypotheses.

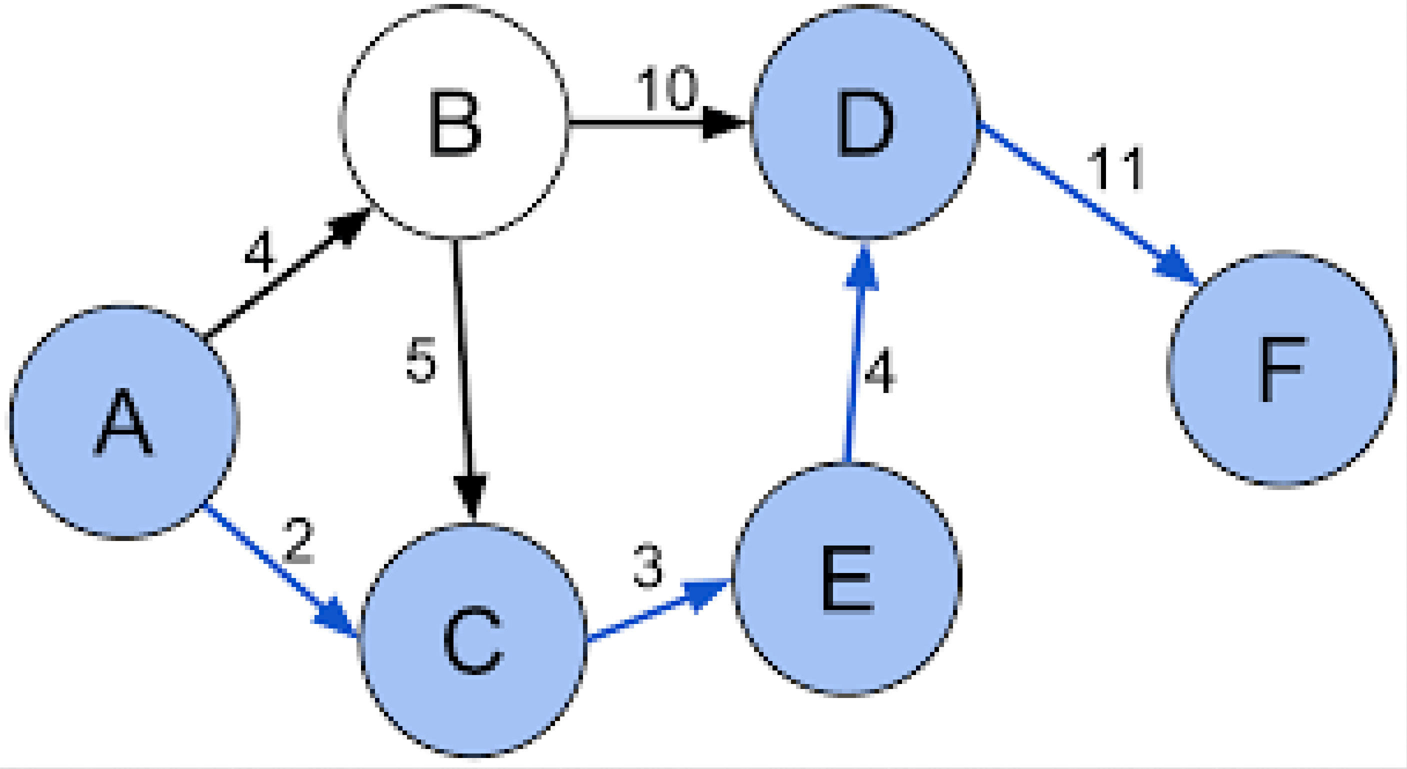
SEARCH



What is shortest path from Node 5(A) to Node 3(B)?

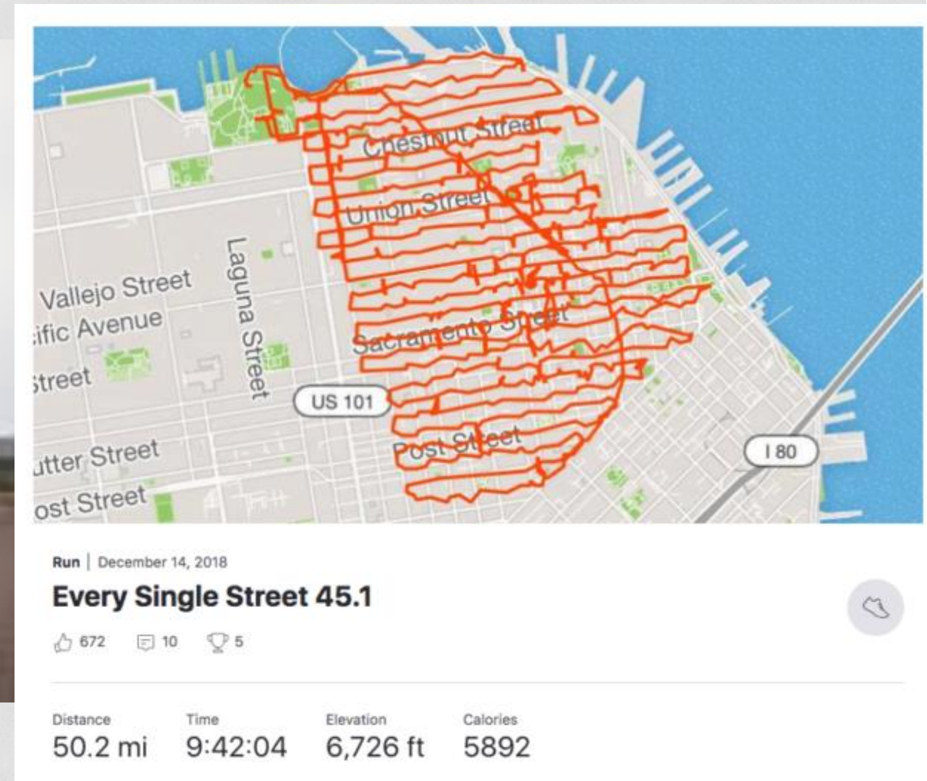
SHORTEST PATH ALGORITHMS (LARGELY SOLVED)

Shortest path from A to B

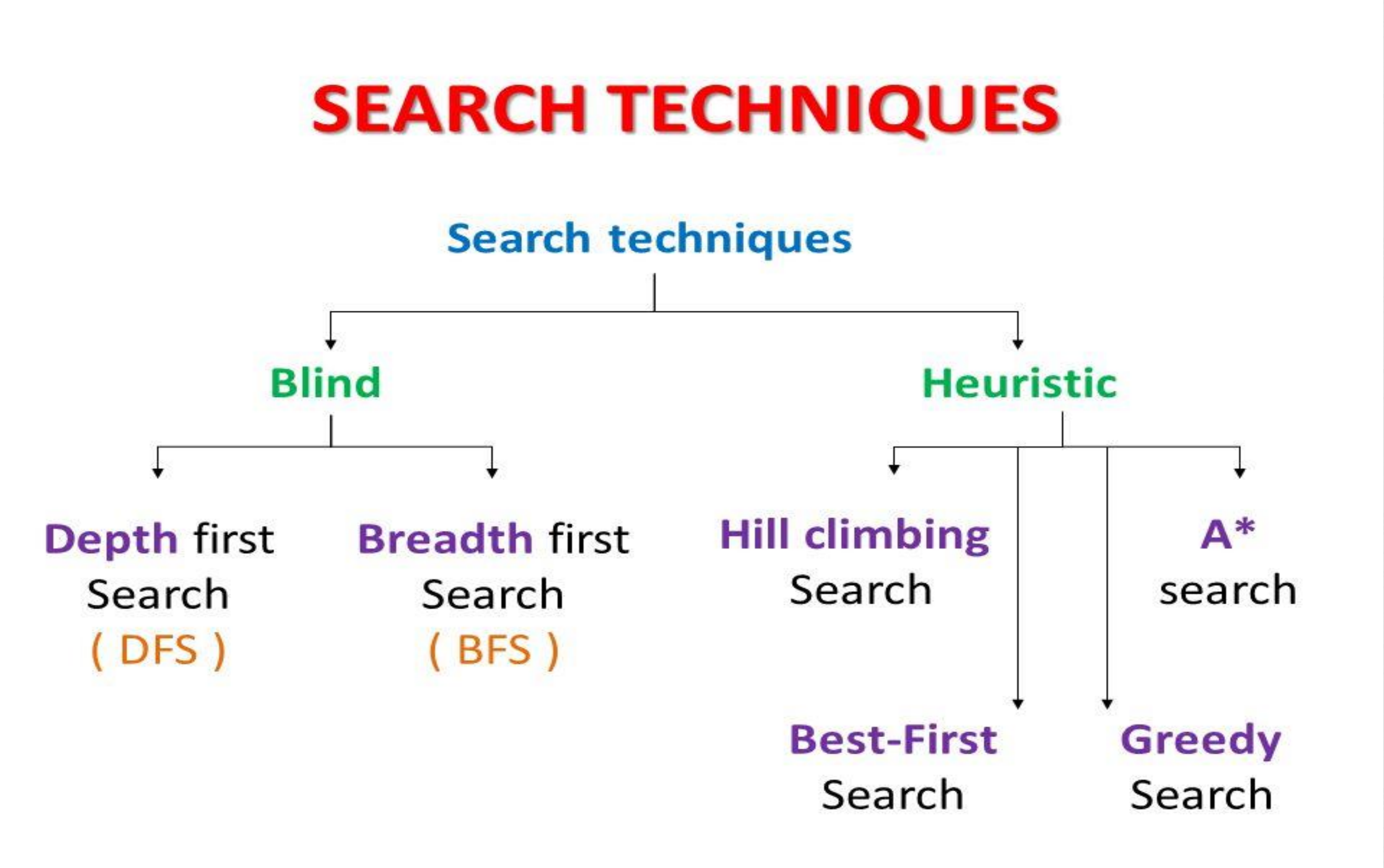


SOME PROBLEMS ARE TOO HARD!

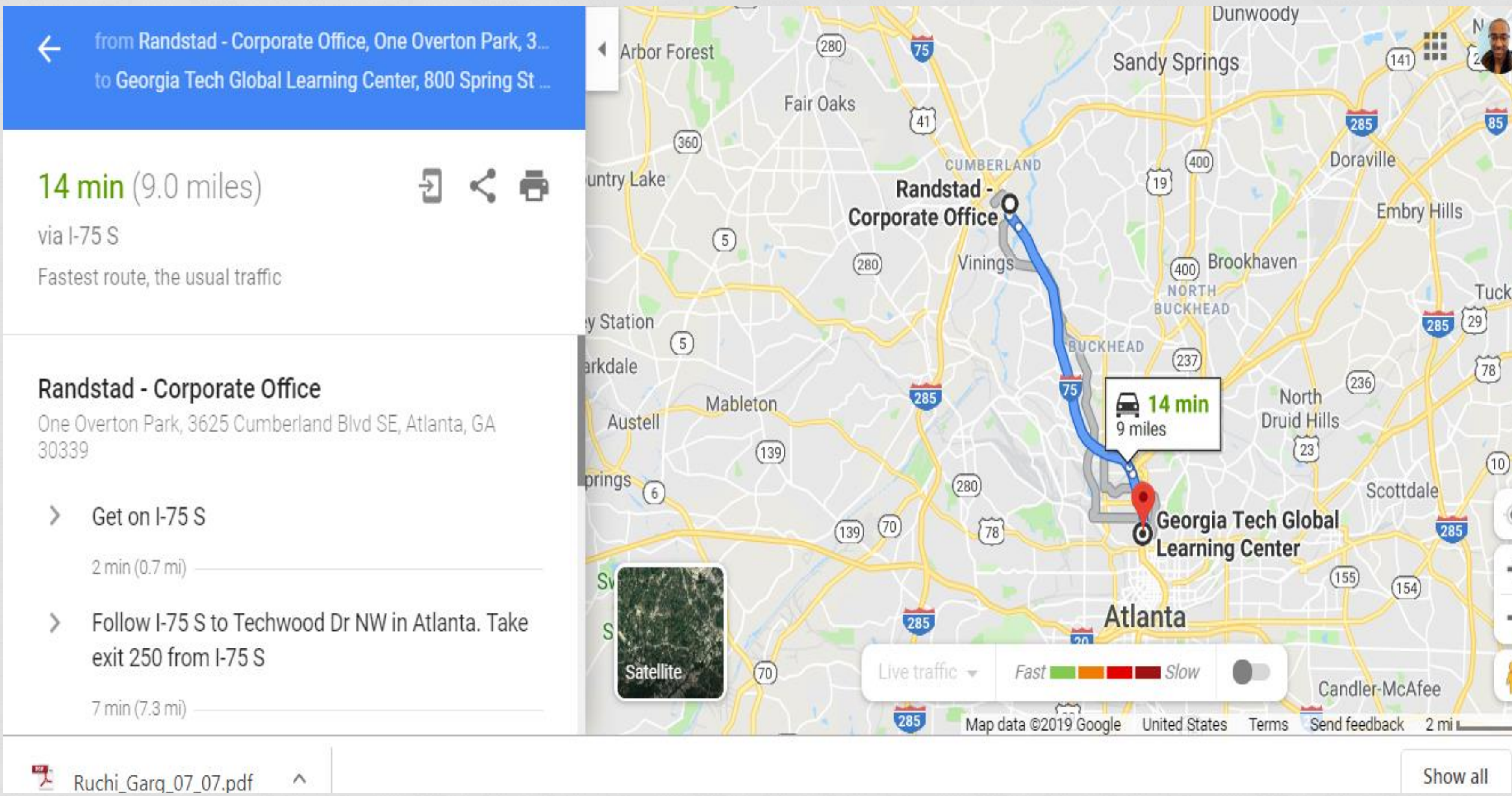
- Shortest path to travel every street on a map



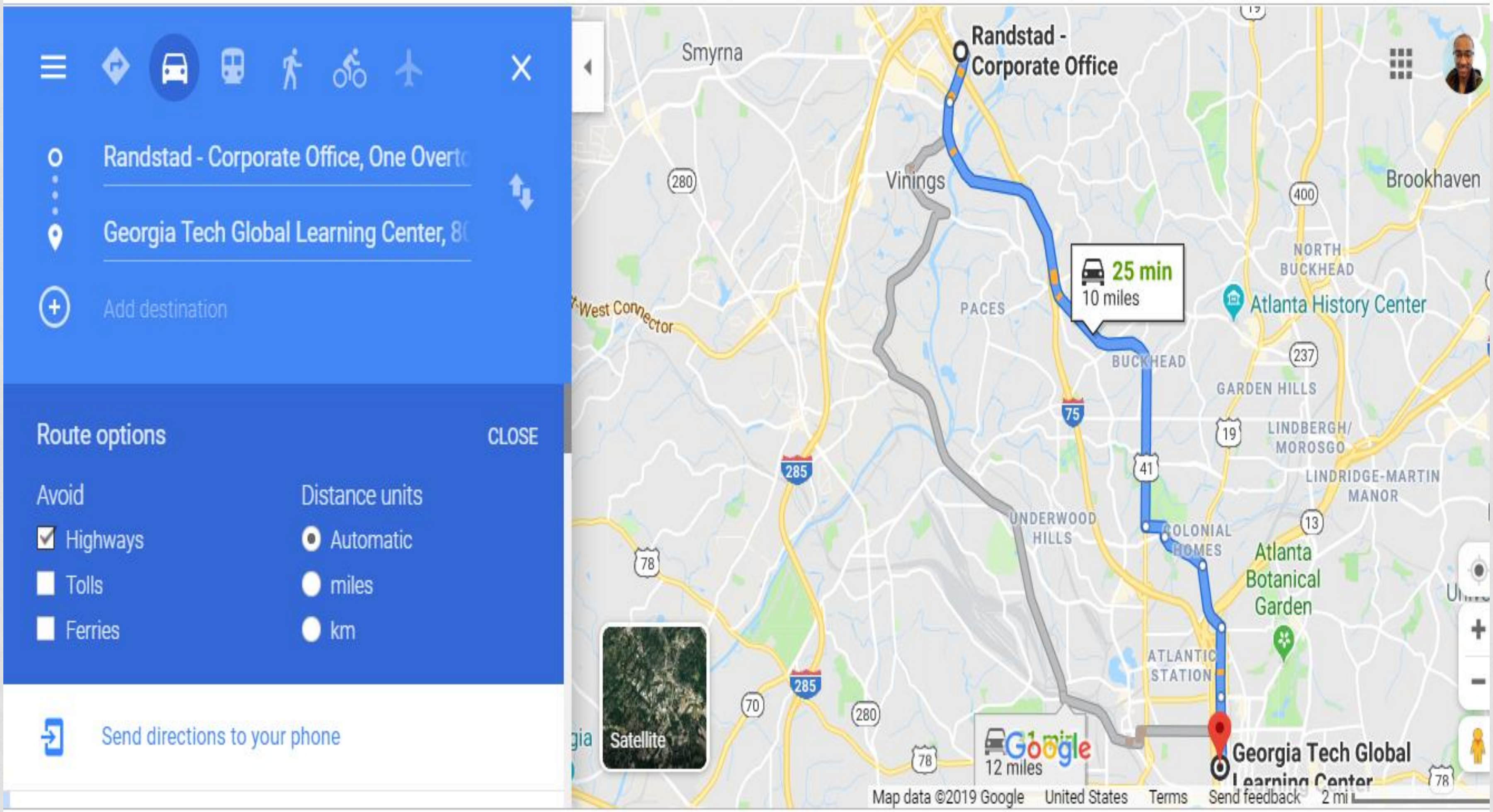
ALGORITHMS FOR SHORTEST PATH PROBLEMS



SHORTEST PATH (GOOGLE MAPS)



SHORTEST PATH (WITH CONSTRAINTS)



ALGORITHMS FOR SORTING

Name	Best	Average	Worst	Memory	Stable	Method	Other notes
Quicksort	$n \log n$ variation is n	$n \log n$	n^2	$\log n$ on average, worst case space complexity is n; Sedgewick variation is $\log n$ worst case.	Typical in-place sort is not stable; stable versions exist.	Partitioning	Quicksort is usually done in-place with $O(\log n)$ stack space.[4][5]
Merge sort	$n \log n$	$n \log n$	$n \log n$	n A hybrid block merge sort is O(1) mem.	Yes	Merging	Highly parallelizable (up to $O(\log n)$ using the Three Hungarians' Algorithm[6] or, more practically, Cole's parallel merge sort) for processing large amounts of data.
In-place merge sort	—	—	$n \log^2 n$ See above, for hybrid, that is $n \log n$	1	Yes	Merging	Can be implemented as a stable sort based on stable in-place merging.[7]
Heapsort	n If all keys are distinct, $n \log n$	$n \log n$	$n \log n$	1	No	Selection	
Insertion sort	n	n^2	n^2	1	Yes	Insertion	$O(n + d)$, in the worst case over sequences that have *d* inversions.

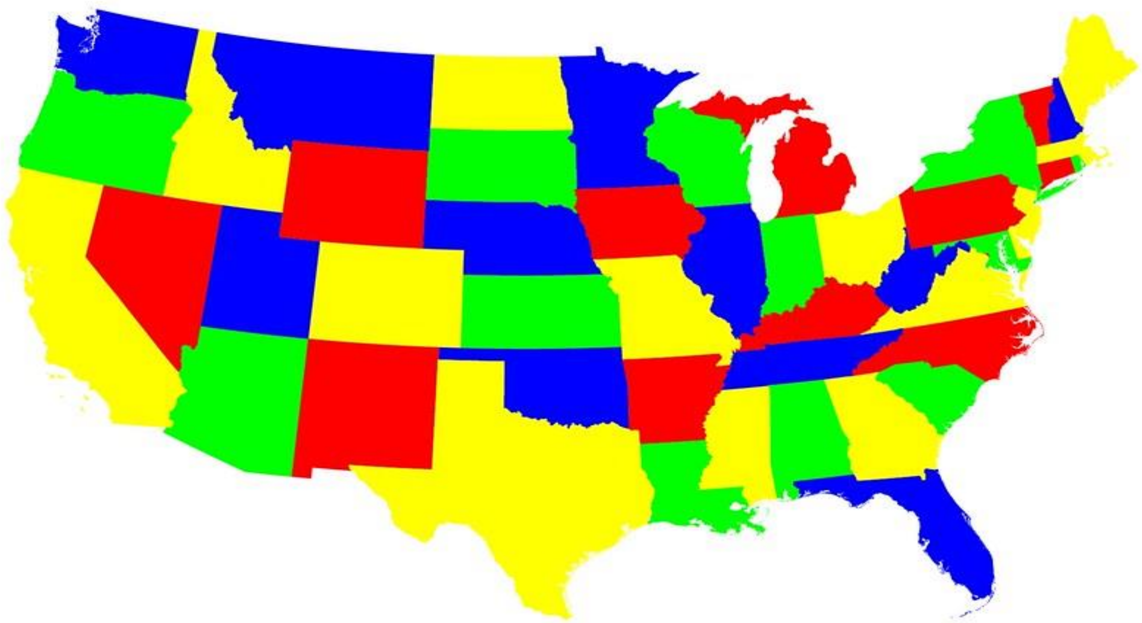
SEARCH - CONSTRAINT SATISFACTION PROBLEMS

- **Constraint satisfaction problems** (**CSPs**) are mathematical questions defined as a set of objects whose state must satisfy a number of constraints or limitations.
- How to find options that work?
  - **Exhaustive search**

MAP COLORING

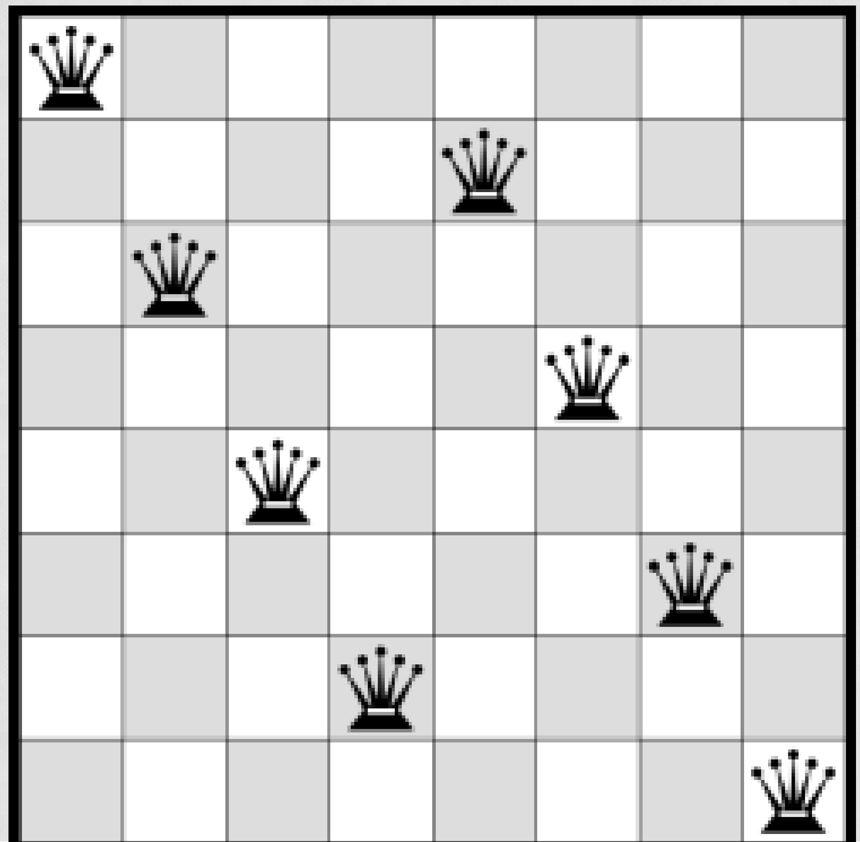
Color states on map where any adjacent state has different color.

Solution – Exhaustive Search



8 QUEENS PROBLEM

- Find configuration where 8 queens on a chess board do not attack each other (i.e., put each other in check)



SUDOKU PUZZLE

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

REAL WORLD CONSTRAINT SATISFACTION PROBLEMS

- Scheduling for calendars (Holiday Schedules)
- Job Recruiting
- Housing buying

PROBABILISTIC REASONING

- Use probability to make decisions
- In a nutshell, there are many options, each with a probability of being the best, simply select the option with the highest probability.
- MAP Decision
  - Maximum A Posteriori (MAP) decision

AXIOMS OF PROBABILITY (REVIEW)

- First Axiom – The probability of an event is a non-negative real number
- Second Axiom – The sum of all probabilities in the event space is 1
- Third Axiom – Sum of the probability of disjoint sets is simply the sum of the disjoint sets.

satisfies

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

https://en.wikipedia.org/wiki/Probability_axioms

NAÏVE BAYES CLASSIFICATION

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

NAÏVE BAYES DETAILS

- P(X|C) - Likelihood – probability distribution given class **(Required)**
  - histograms
- P(C) - Prior – Probability of the Class (usually based on Frequency) – **(Optional)**
  - Banana = 100 – Prior Probability = 1/5
  - Apple = 300 – Prior Probability = 3/5
  - Peach = 100 – Prior Probability = 1/5
- P(X) – Probability of the data - **ignore**

Derivation of Naive Bayes rule

We want to find the class that is most likely given the document:

$$c_{\text{map}} = \underset{c \in \mathbb{C}}{\arg\max}\ P(c|d)$$

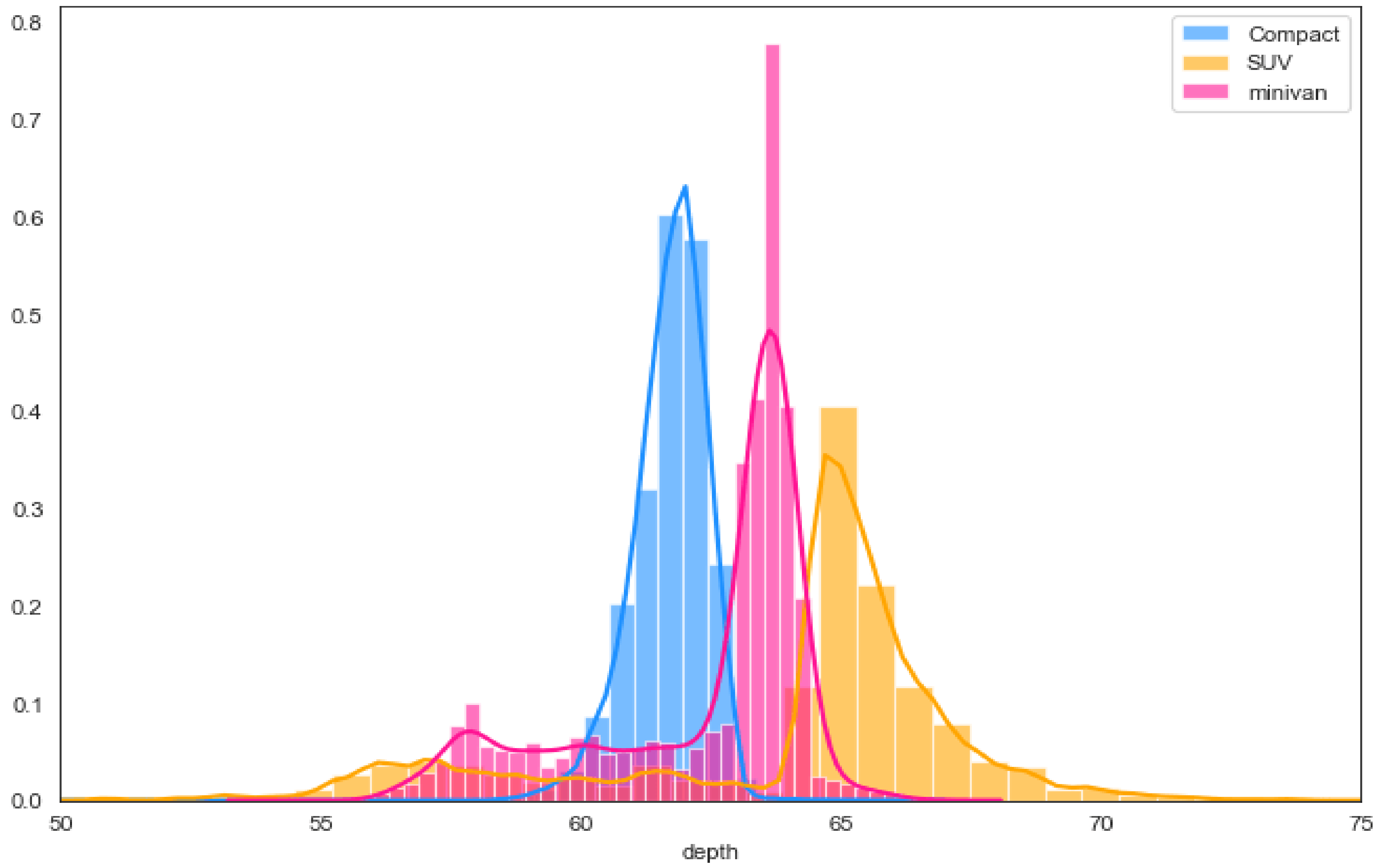
Apply Bayes rule $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$:

$$c_{\text{map}} = \underset{c \in \mathbb{C}}{\arg\max}\ \frac{P(d|c)P(c)}{P(d)}$$

Drop denominator since P(d) is the same for all classes:

$$c_{\text{map}} = \underset{c \in \mathbb{C}}{\arg\max}\ P(d|c)P(c)$$

NAÏVE-BAYES HISTOGRAM-BASED EXAMPLE



MACHINE LEARNING TECHNIQUES

- Unsupervised
  - Clustering
- Supervised
  - Regression
  - Classification

COMPONENTS OF A MACHINE LEARNING TECHNIQUE

- Input Data – Features (X) and Target Y (Optional)
- Objective to optimize
  - Parameters
  - Maximizing or minimizing something
  - Model Type - Classification, Clustering, Regression
- Optimization Algorithm
- Validation
  - Metrics – Accuracy, mean square error, Likelihood, F-score, Plots, Qualitative
  - Evaluation Methodology to use for supervised learning – Cross validation

LINEAR REGRESSION

- Model Type – Regression
- Objective – Minimize mean square error
- Optimization algorithm – Ordinary Least Squares, gradient descent

LOGISTIC REGRESSION

- Model Type – Classification
- Objective – Logit
- Optimization algorithm – gradient descent, newton method

NEURAL NETWORKS

- Model Type – Classification/Regression
- Objective – Defined by User
- Optimization Algorithm – Back Propagation