

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
0- Pilot

Examples:



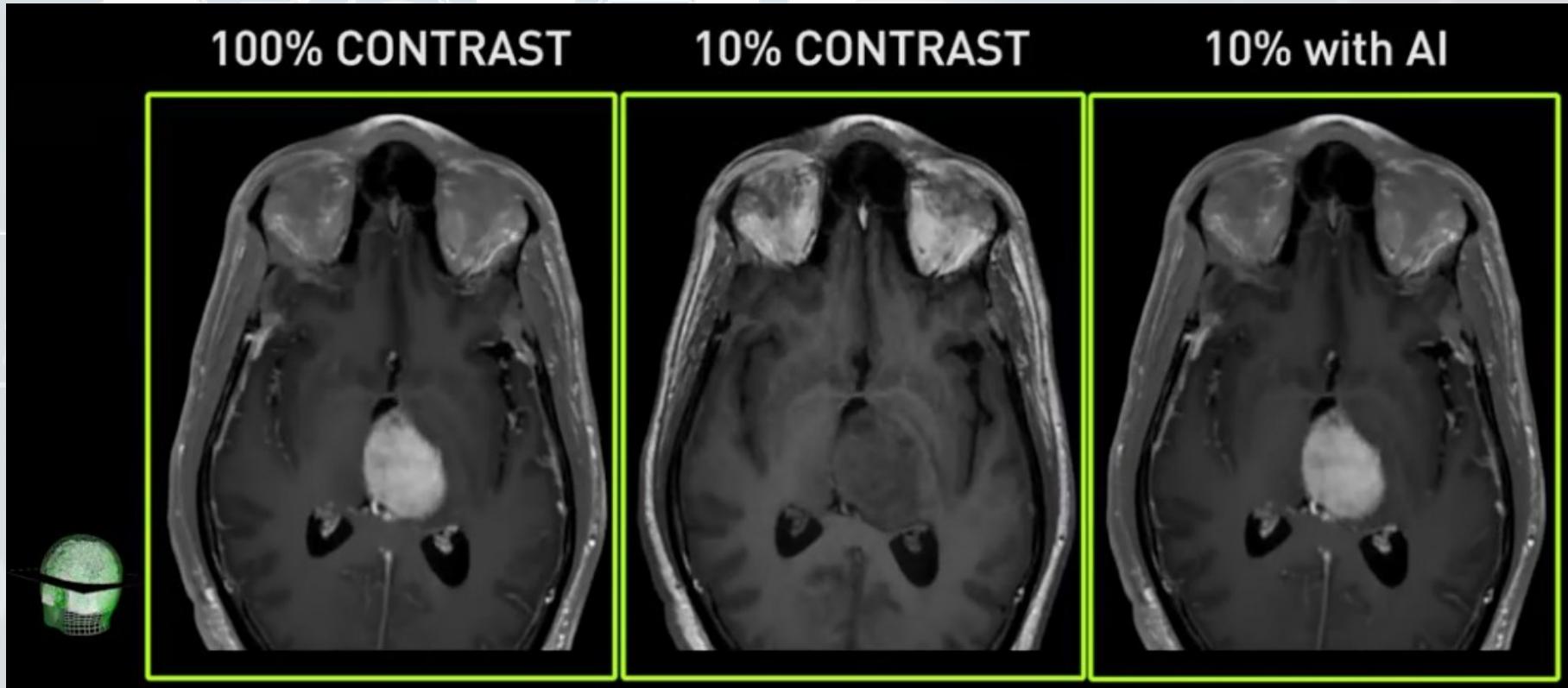
- 10 Likes for the computer to outperform a work colleague
- 70 to do better than a friend
- 150 to outscore a family member
- 227 to almost predict you as good as your spouse (if you are married!)

Computer-based personality judgments are more accurate than those made by humans

Examples:



Examples:



Examples:

In the US dataset, the software performed significantly better than human experts, producing 5.7 percent fewer false positive diagnoses. Even more impressively, the system recorded 9.4 percent fewer false negatives, suggesting it picked up several breast cancers that human experts missed.



Course journey (big picture):

- Introduction 
- What is in the beginner course? 
- Python 
- Statistics 
- ML and AI 
- Mini-courses / challenges 

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
0.5- Ground rules



GROUND RULES



EVERYTHING

I say or show here
is shared with you!!!



DON'T get disappointed!



DON'T memorize things!



DON'T underestimate yourself!

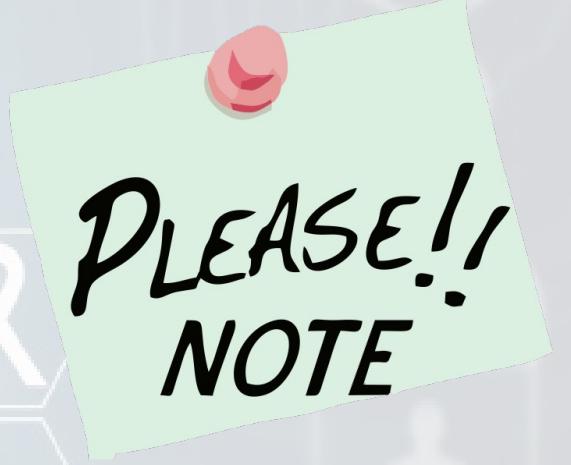


DON'T overestimate AI/engineers!



DON'T EVER FORGET Google~~X~~ING!!!





PLEASE!!
NOTE

The main stream of the course is very easy!

The exercises are designed to make us (you and me) sure that you don't miss important points!

There are some auxiliary sessions for who is interested.

They are not supposed to affect the course connectivity so .. simply skip them if you like!

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr

1- Introduction:

Have we ever seen data science!?

Some cool examples

Spotify, Google translate, Siri, Cortana,
Google digital assistants, Pinterest,
Instagram, Facebook and LinkedIn, Uber,
Tapsi, Gmail, Waze



Google for “Google's DeepMind AI Just Taught Itself To Walk”

Google for “Looking to Listen: Audio-Visual Speech Separation”

Google for “AI brings Mona Lisa to life”

Google Assistant

Fast Facts

Famous Data Scientist



Larry Page
CEO of Google

Job Opportunities

15,000%

increase in job postings for data scientists between 2011 & 2012.

Majors



physics



applied maths



social sciences



statistics



analytics



computer science



marketing

\$80K

average starting salary

\$120K

average data science salary

\$250K

data science team manager

\$400K

highest paid data scientist

thanks
for
watching!

MEDICAL DATA SCIENCE

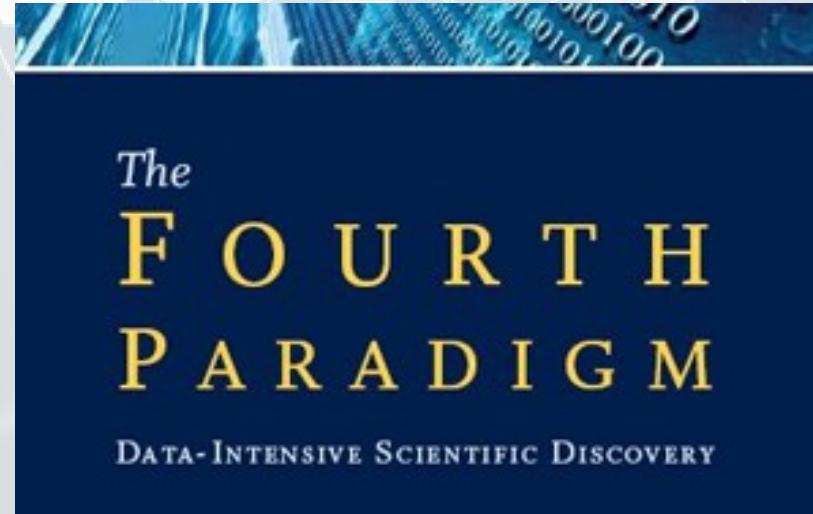
An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr

2- Introduction:

Data science story! - part 1

Science Story





Empirical evidence

$$\Psi(x) = \frac{1}{\sqrt{2\pi}}(A e^{ix} + A^* e^{-ix}) \quad x < 0$$

$$K = \sqrt{2mE/\hbar^2} \quad R_p = \frac{1}{2}R_g + \Delta g_p = \frac{8\pi G}{c^3} T_p$$

$$H = \frac{P_p P_e}{2m} + V(r) \quad P = -i\hbar\nabla$$

$$H|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle \quad \delta(k_1 + k_2) = \frac{\delta(k_1)}{k_1}$$

$$I = \int e^{-ax^2/2} dx = \sqrt{\frac{2\pi}{a}}$$

$$A_{ij} = \frac{8\pi\hbar v^3}{c^3} B_{ij}$$

$$S_f = \langle f | S | f \rangle$$

$$G_m = R_p - \frac{1}{2}R_g g_p = \frac{2\pi r^2}{c^4} T_p$$

$$\sigma = \frac{2\pi r^2}{T_p^2} (1 - e^{-T_p})$$

$$S_B = \frac{k_B 4\pi G}{\hbar c} M^2$$

$$S = \frac{1}{2k} \int R \sqrt{-g_d} d^4x$$

$$L = \text{tr} \left\{ \frac{i}{\hbar} F_{ij} F^{ij} - i \lambda \Gamma^i D_i \lambda \right\}$$

$$E = mc^2 \quad E^2 = (pc)^2 + (mc^2)^2$$

$$r = \frac{\theta}{2\pi} + \frac{4\pi}{9t}$$

$$P = \hbar k = \frac{\hbar v}{c} = \frac{\hbar}{\lambda}$$

$$S = \frac{1}{2} \int d^4x \left(R + \frac{R^2}{6M^2} \right)$$

$$\Omega_p = 10$$

$$\int V = e^{-\mu} \int_0^s V(X_{t+r}) dr - \int_0^s (\lambda_s)^{\alpha} \frac{\partial V}{\partial X} dW$$

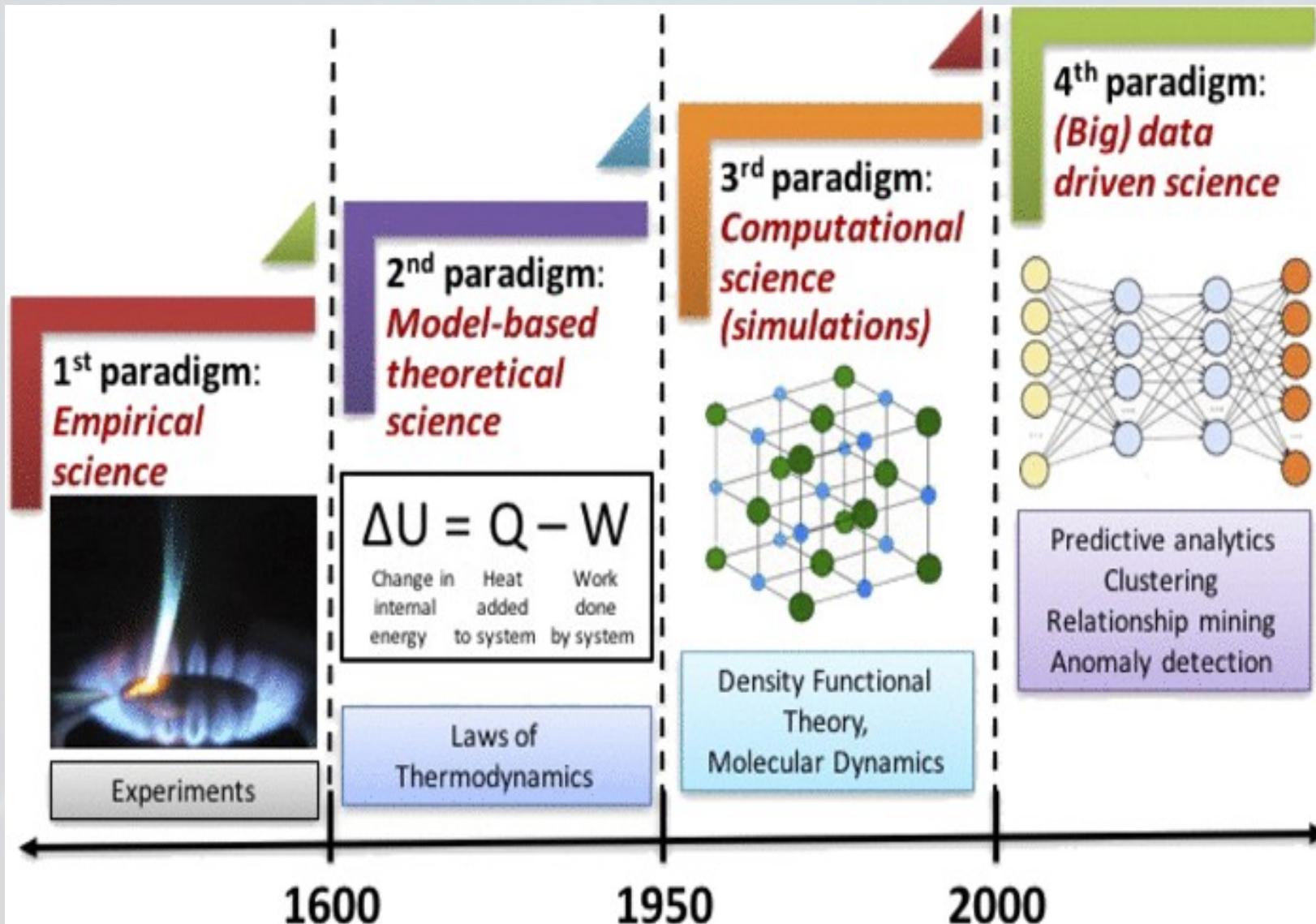
Scientific theory



Computational science

BIG DATA

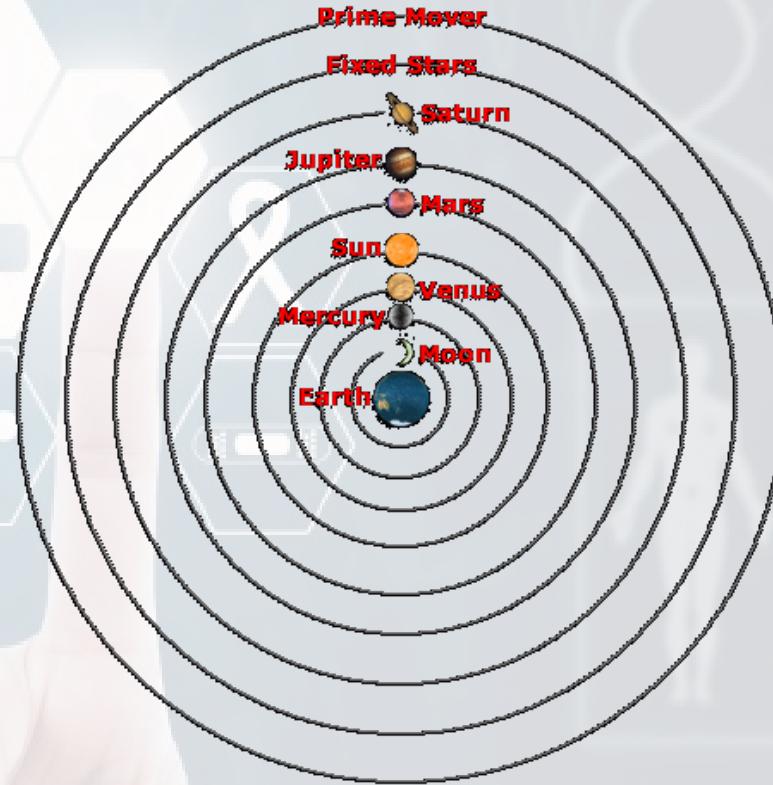
Data science



Empirical evidence



Empirical evidence



Empirical evidence



thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr

3- Introduction:

Data science story! - part 2

1st paradigm:
Empirical science



Experiments

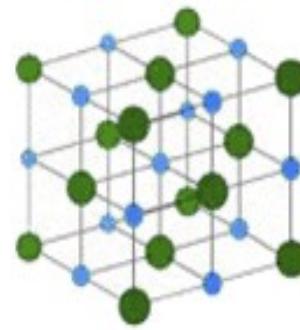
2nd paradigm:
Model-based theoretical science

$$\Delta U = Q - W$$

Change in internal energy Heat added to system Work done by system

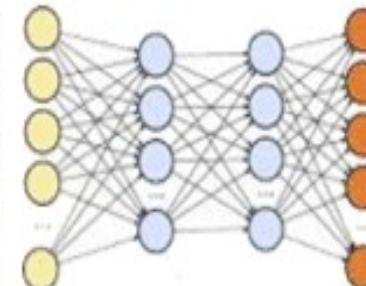
Laws of Thermodynamics

3rd paradigm:
Computational science (simulations)



Density Functional Theory,
Molecular Dynamics

4th paradigm:
(Big) data driven science



Predictive analytics
Clustering
Relationship mining
Anomaly detection

1600

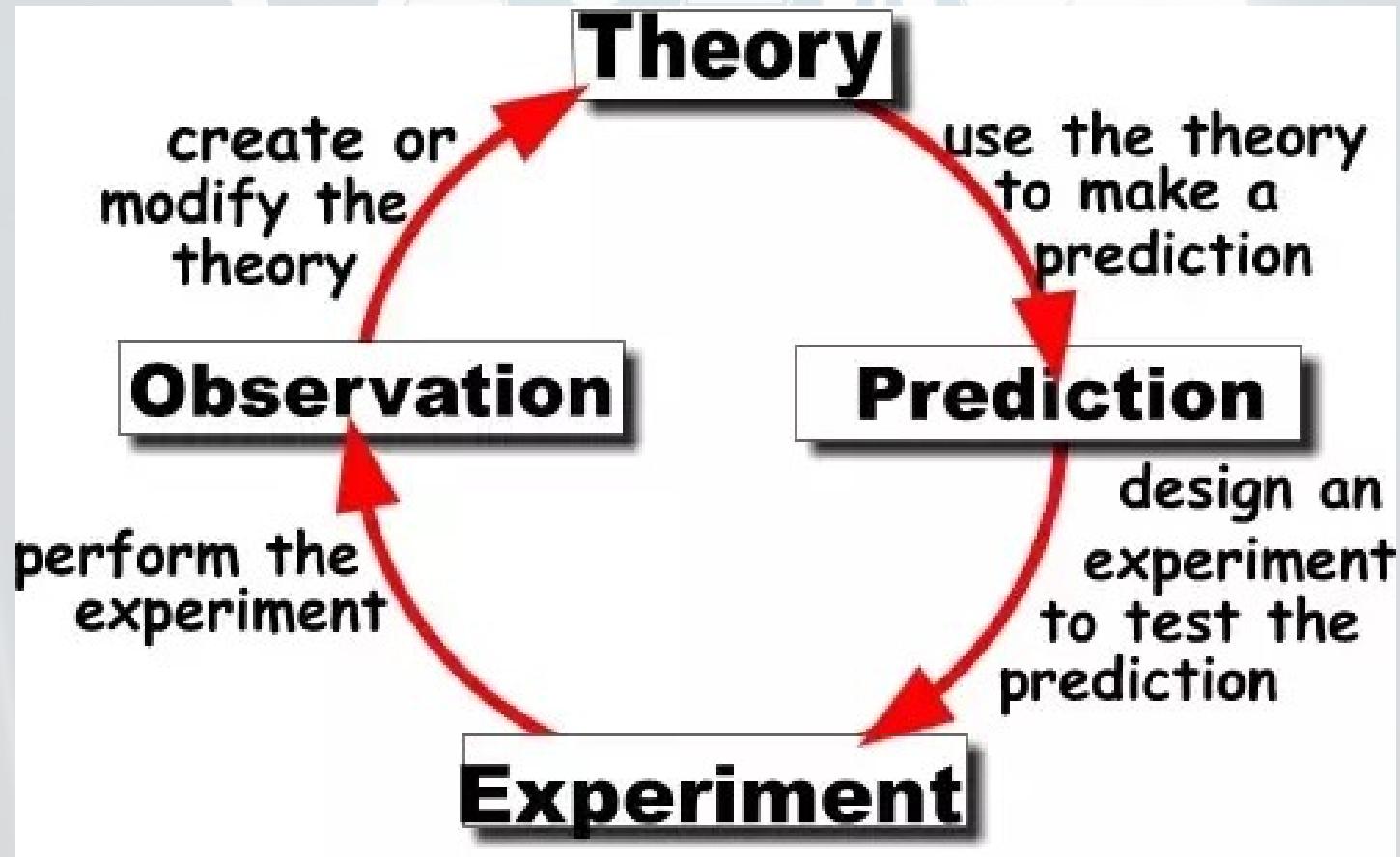
1950

2000

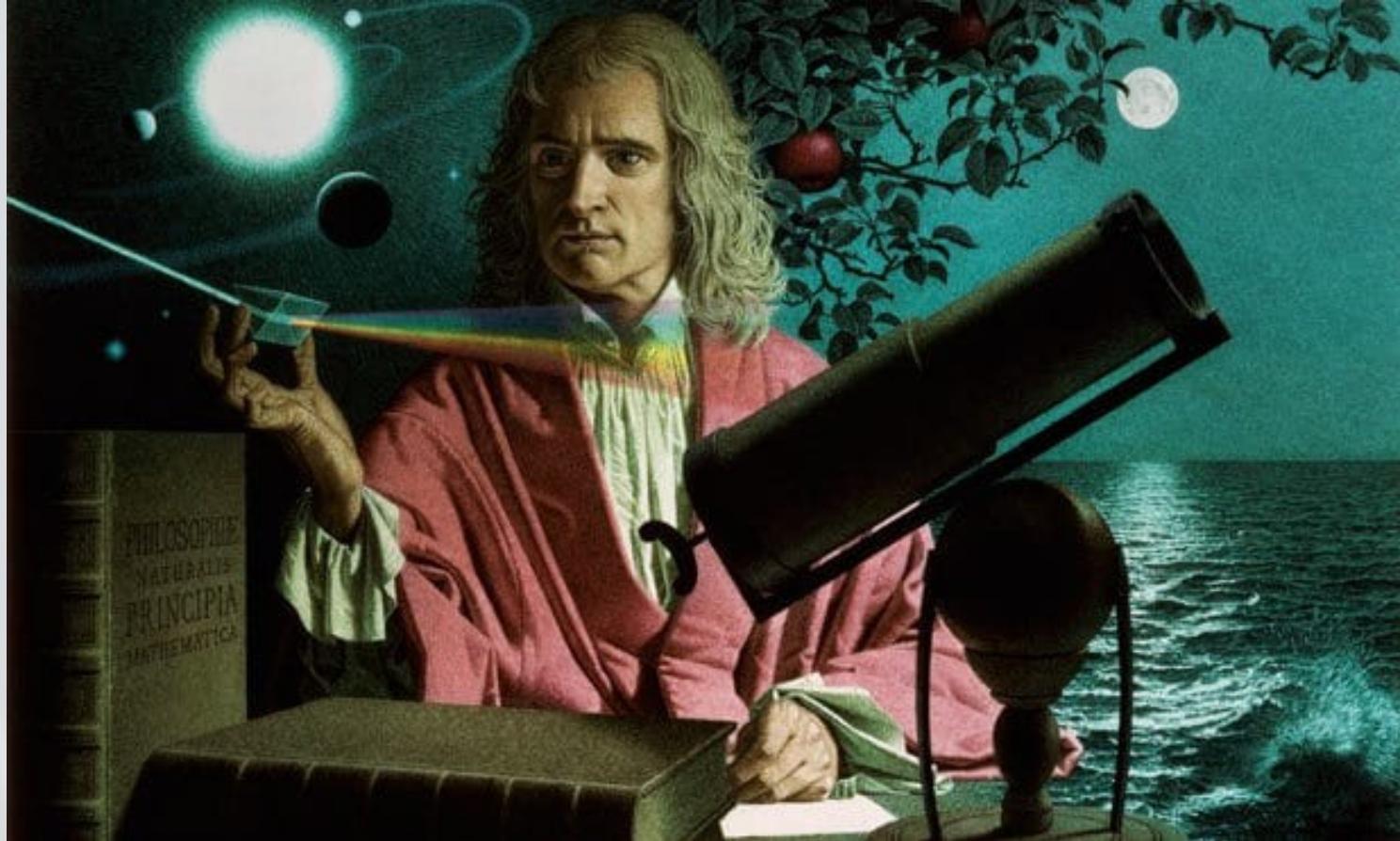
Scientific theory



Scientific theory



Scientific theory



thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr

4- Introduction:

Data science story! - part 3

1st paradigm:
Empirical science



Experiments

2nd paradigm:
Model-based theoretical science

$$\Delta U = Q - W$$

Change in internal energy Heat added to system Work done by system

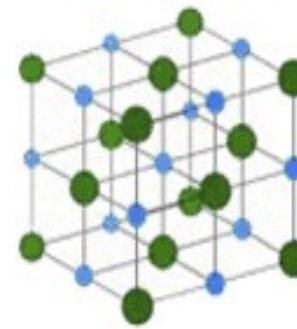
Laws of Thermodynamics

1600

1950

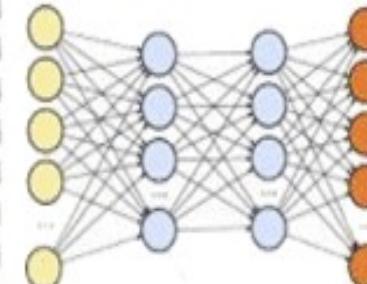
2000

3rd paradigm:
Computational science (simulations)



Density Functional Theory,
Molecular Dynamics

4th paradigm:
(Big) data driven science



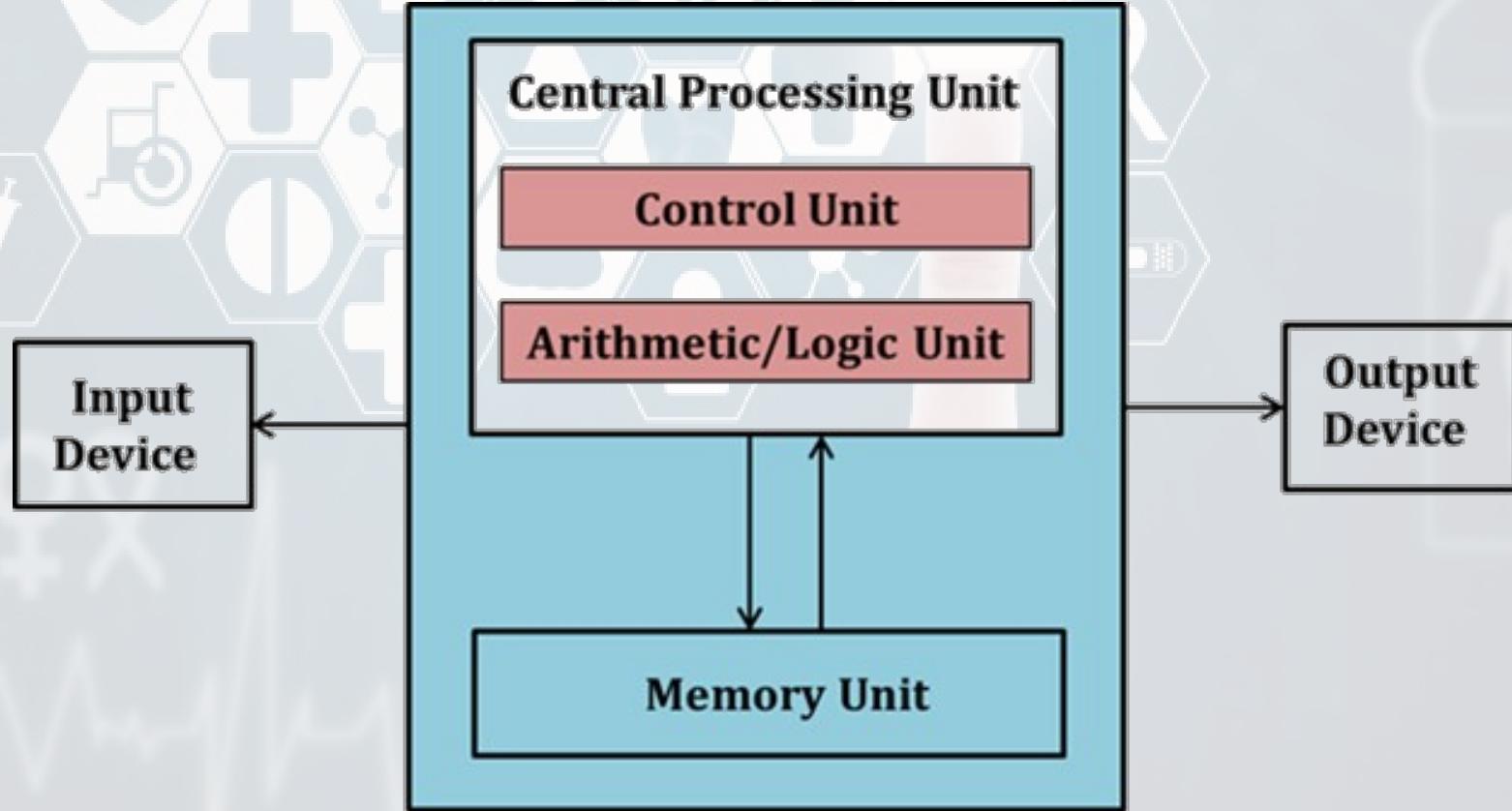
Predictive analytics
Clustering
Relationship mining
Anomaly detection

Computational sciences

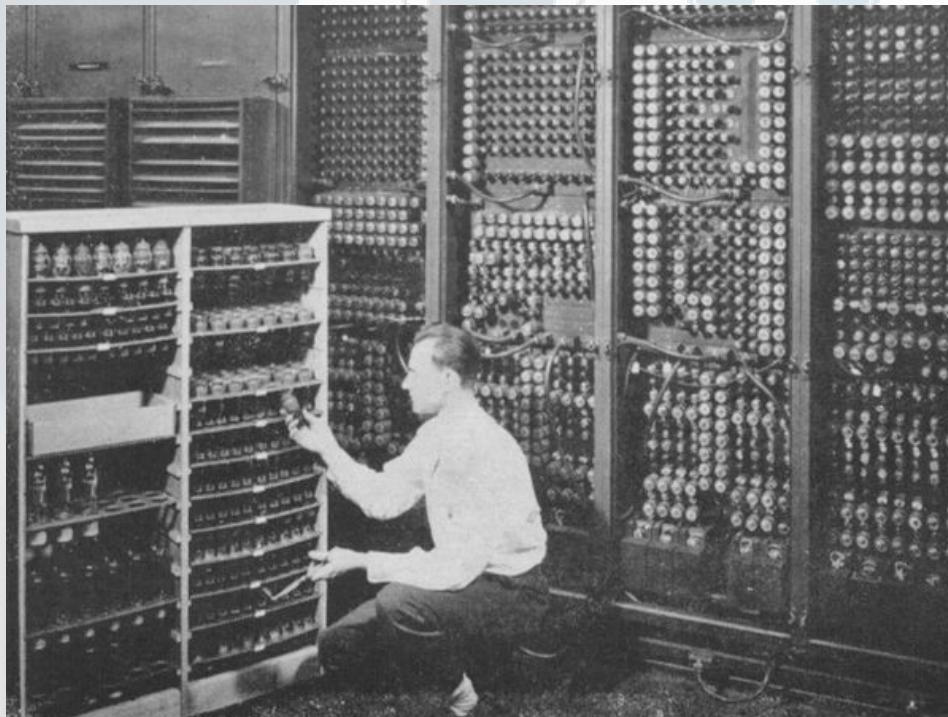
Computer 1613-1950



Computational sciences



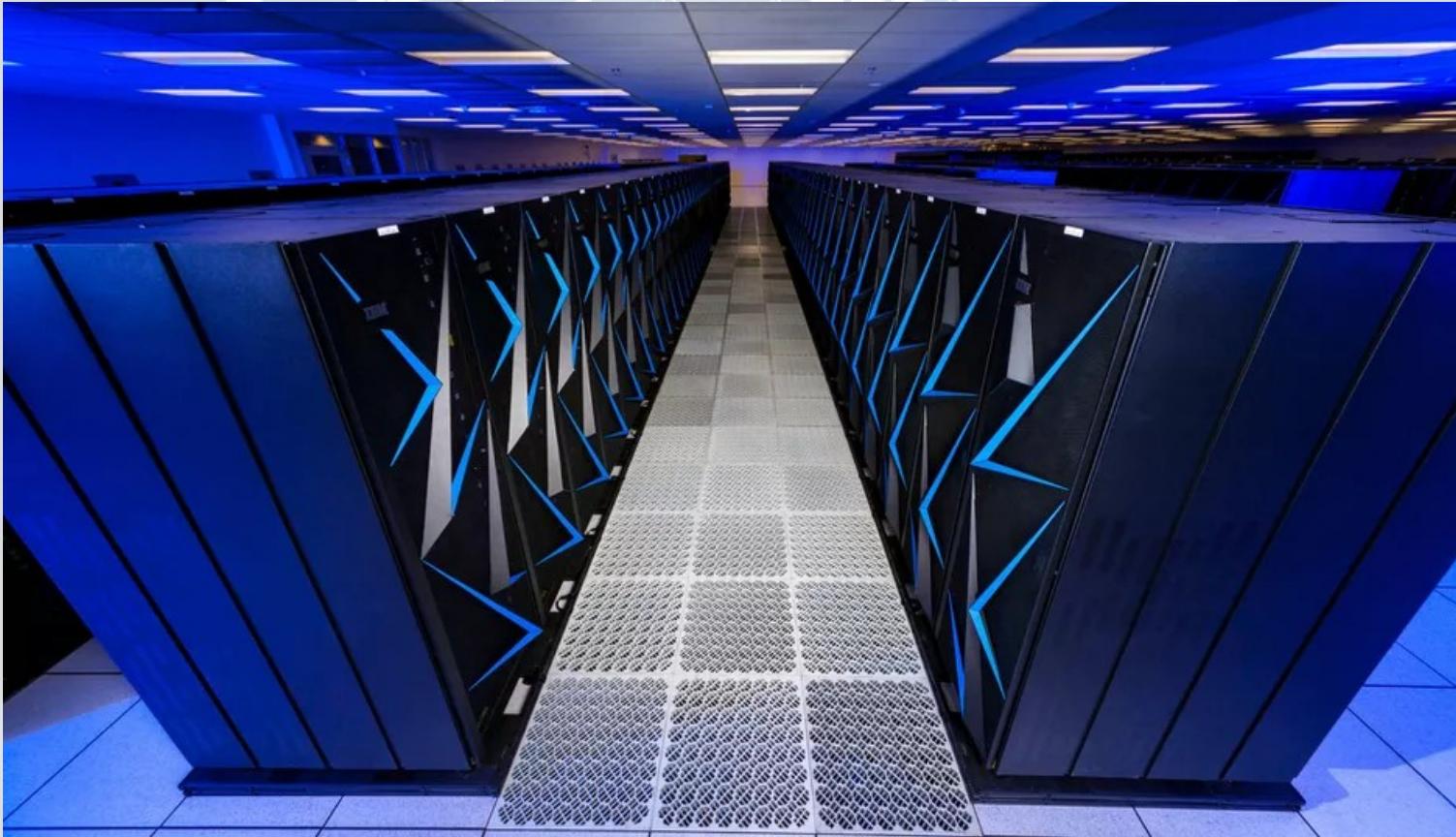
عصر کامپیوٹر



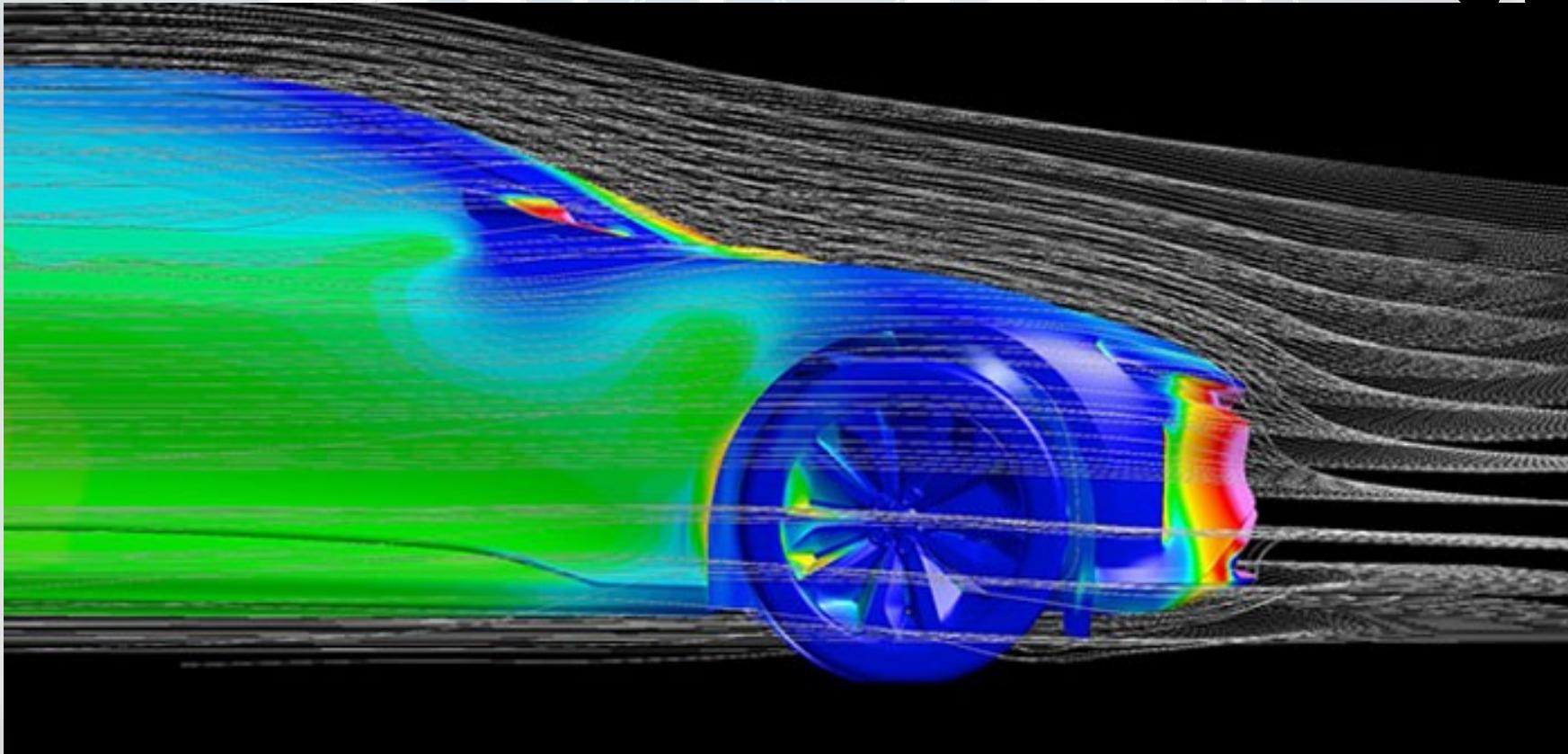
عصر کامپیوٹر



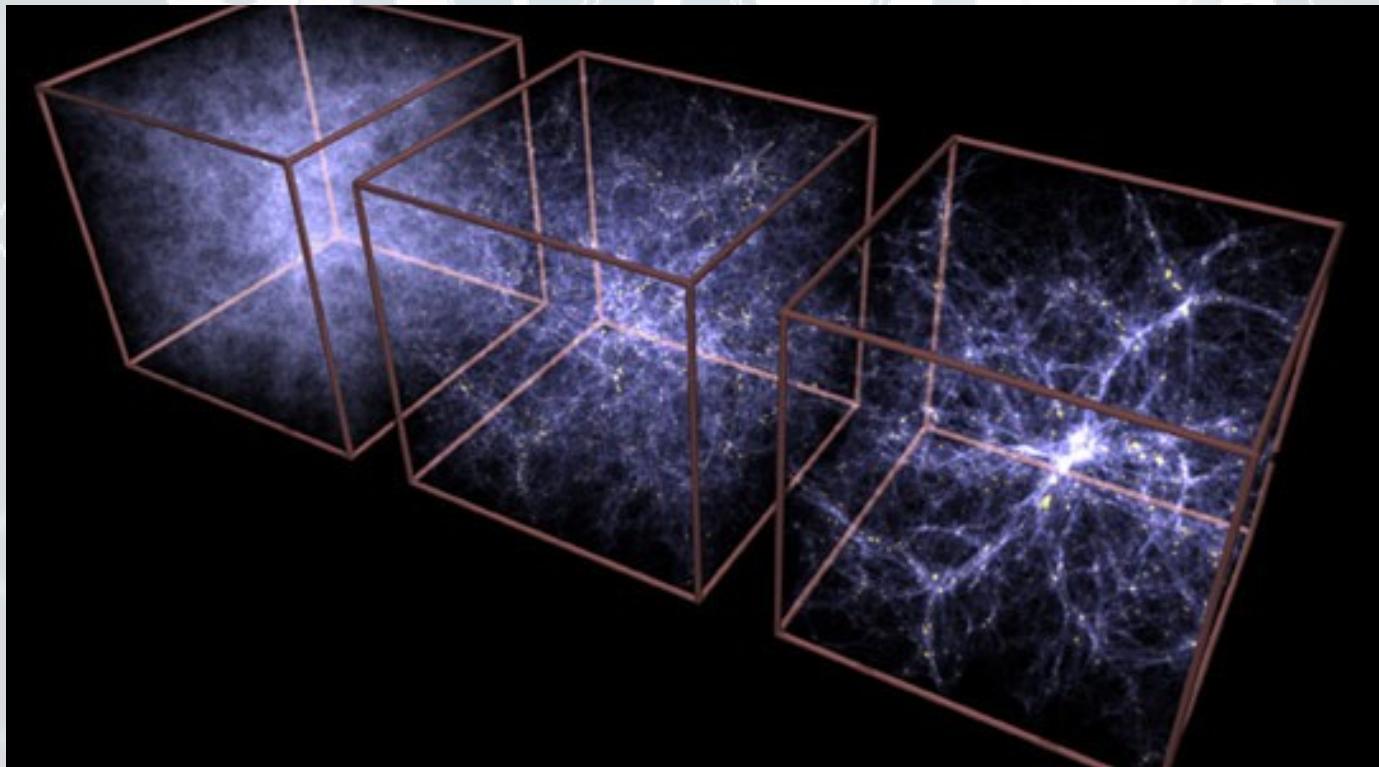
عصر کامپیوٹر



عصر کامپیوٹر



عصر کامپیوٽر



thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
5- Introduction:
Data science story! - part 4

1st paradigm:
Empirical science



Experiments

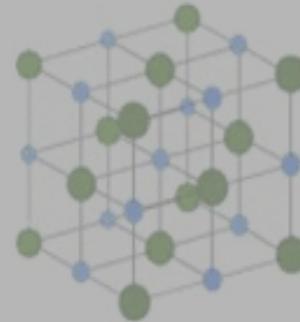
2nd paradigm:
Model-based theoretical science

$$\Delta U = Q - W$$

Change in internal energy Heat added to system Work done by system

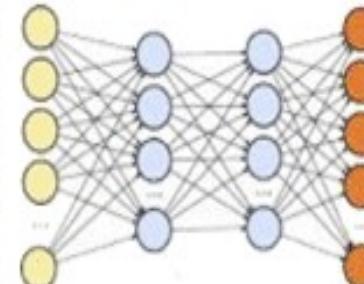
Laws of Thermodynamics

3rd paradigm:
Computational science (simulations)



Density Functional Theory,
Molecular Dynamics

4th paradigm:
(Big) data driven science



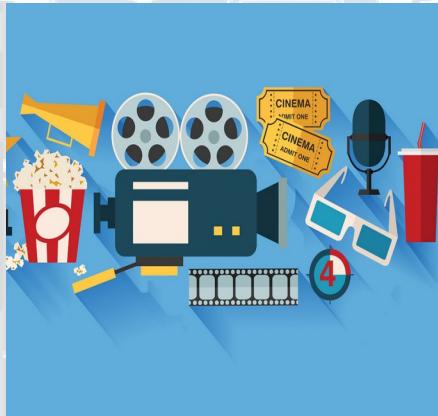
Predictive analytics
Clustering
Relationship mining
Anomaly detection

1600

1950

2000

Data science



Data science



Agriculture



Climate



Consumer



Ecosystems



Education



Energy



Finance



Health



Local
Government



Manufacturing



Maritime



Ocean

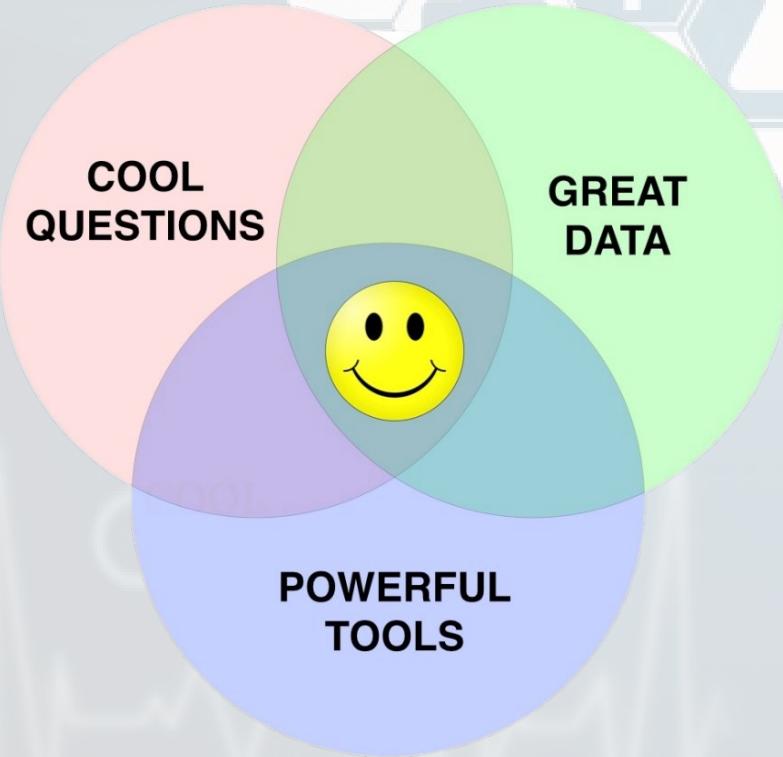


Public Safety



Science &
Research

Data science



Data science

- Hypothesis-Driven:
What kind of data do we need to help solve a problem?
- Data-Driven:
What interesting problems can be solved with this data!?



"Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway." – Geoffrey Moore

"Hiding within those mounds of data is knowledge that could change the life of a patient, or change the world." – Atul Butte

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
6- Introduction:
Some DS cool examples

Some cool examples

Spotify, Google translate, Siri, Cortana,
Google digital assistants, Pinterest,
Instagram, Facebook and LinkedIn, Uber,
Tapsi, Gmail, Waze



Google for “Google's DeepMind AI Just Taught Itself To Walk”

Google for “Looking to Listen: Audio-Visual Speech Separation”

Google for “AI brings Mona Lisa to life”

Creativity!



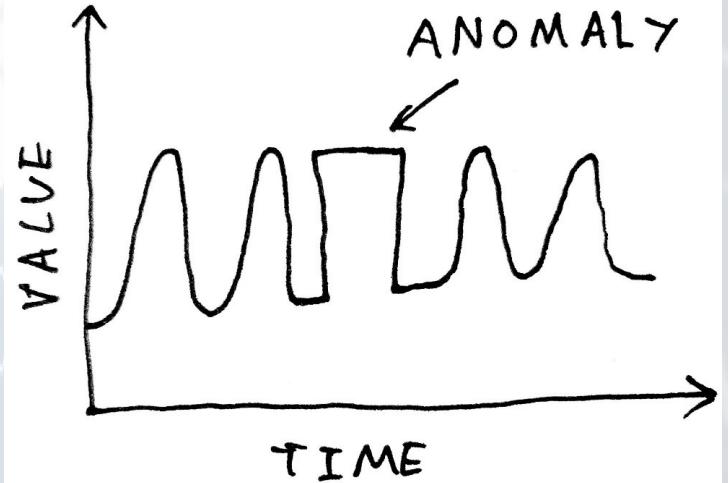
Creative adversarial networks: generating art!

Creativity!



Google for “These People Do Not Exist!”

Discovery!



Discovery!

Table 1 Major discoveries made by the Hubble Space Telescope (*HST*). Of the *HST*'s “top ten” discoveries (as ranked by National Geographic magazine), only one was a key project used in the *HST* funding proposal (Lallo 2012). A further four projects were planned in advance by individual scientists but not listed as key projects in the *HST* proposal. Half the “top ten” *HST* discoveries were unplanned, including two of the three most cited discoveries, and including the only *HST* discovery (Dark Energy) to win a Nobel prize. This Table was previously published by Norris et al. (2015).

Project	Key Project?	Planned?	Nat Geo top ten?	Highly cited?	Nobel Prize?
Use cepheids to improve value of H_0	✓	✓	✓	✓	
UV spectroscopy of ig medium	✓	✓			
Medium-deep survey	✓	✓			
Image quasar host galaxies		✓	✓		
Measure SMBH masses		✓	✓		
Exoplanet atmospheres		✓	✓		
Planetary Nebulae		✓	✓		
Discover Dark Energy			✓	✓	✓
Comet Shoemaker-Levy			✓		
Deep fields (HDF, HDFS, GOODS, FF, etc)			✓	✓	
Proplyds in Orion			✓		
GRB Hosts			✓		

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
7- Introduction:
BIG data?

How BIG?

- New telescopes collect today 50 times the info they collected 5 years ago.
- Google process 24 PB per day = US library of congress X1000
- Facebook updates 10M photos per hour and 38B like per day.
- YouTube adds one hour of video every second



BIG data



One 4K (Ultra HD) movie is about 180 GB



thanks
for
watching!

MEDICAL DATA SCIENCE

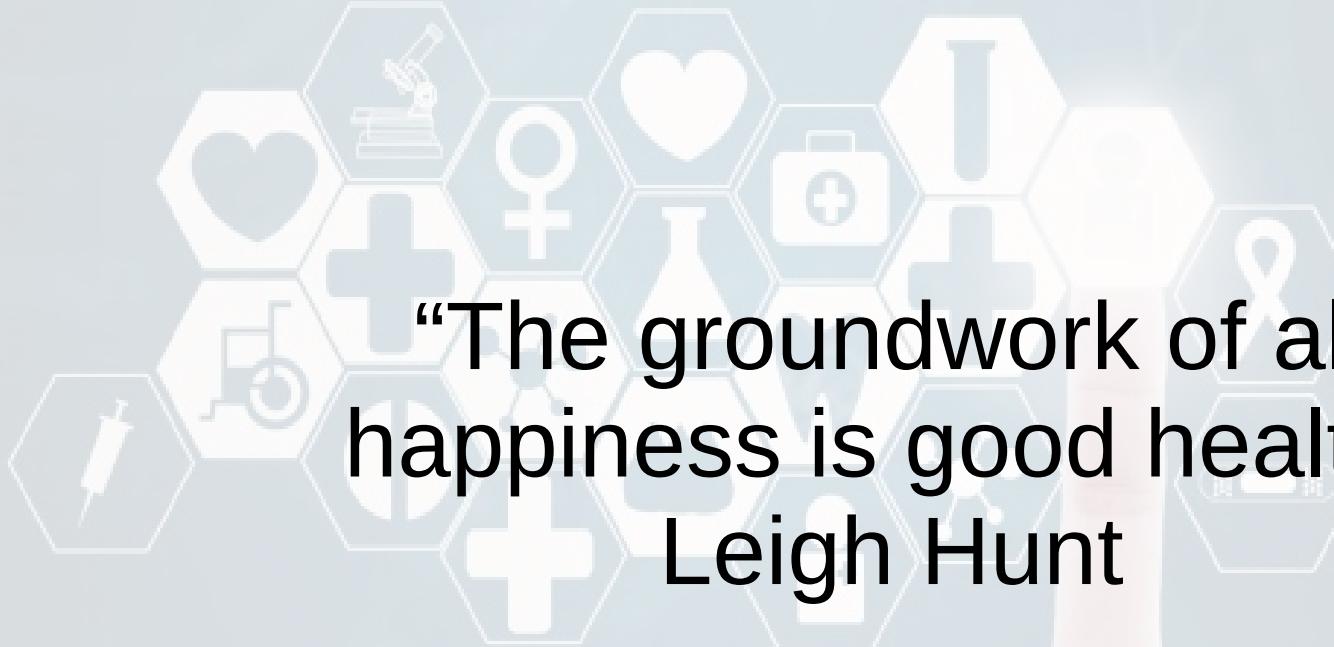
An introduction to:
data science applications in healthcare

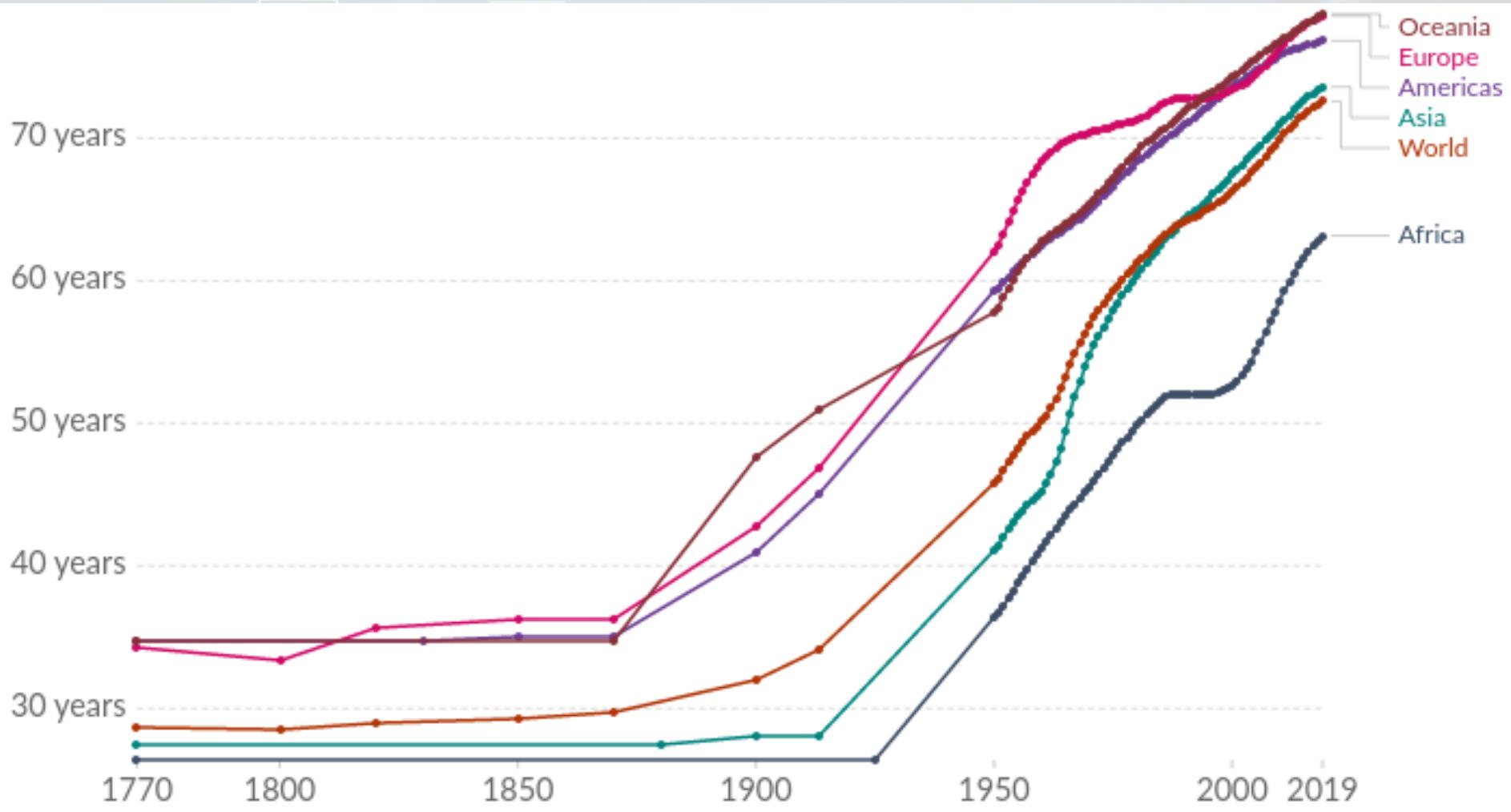
By: Alireza Vafaei Sadr
8- Introduction:

Intelligent Medicine and healthcare system?



“The groundwork of all
happiness is good health”
Leigh Hunt







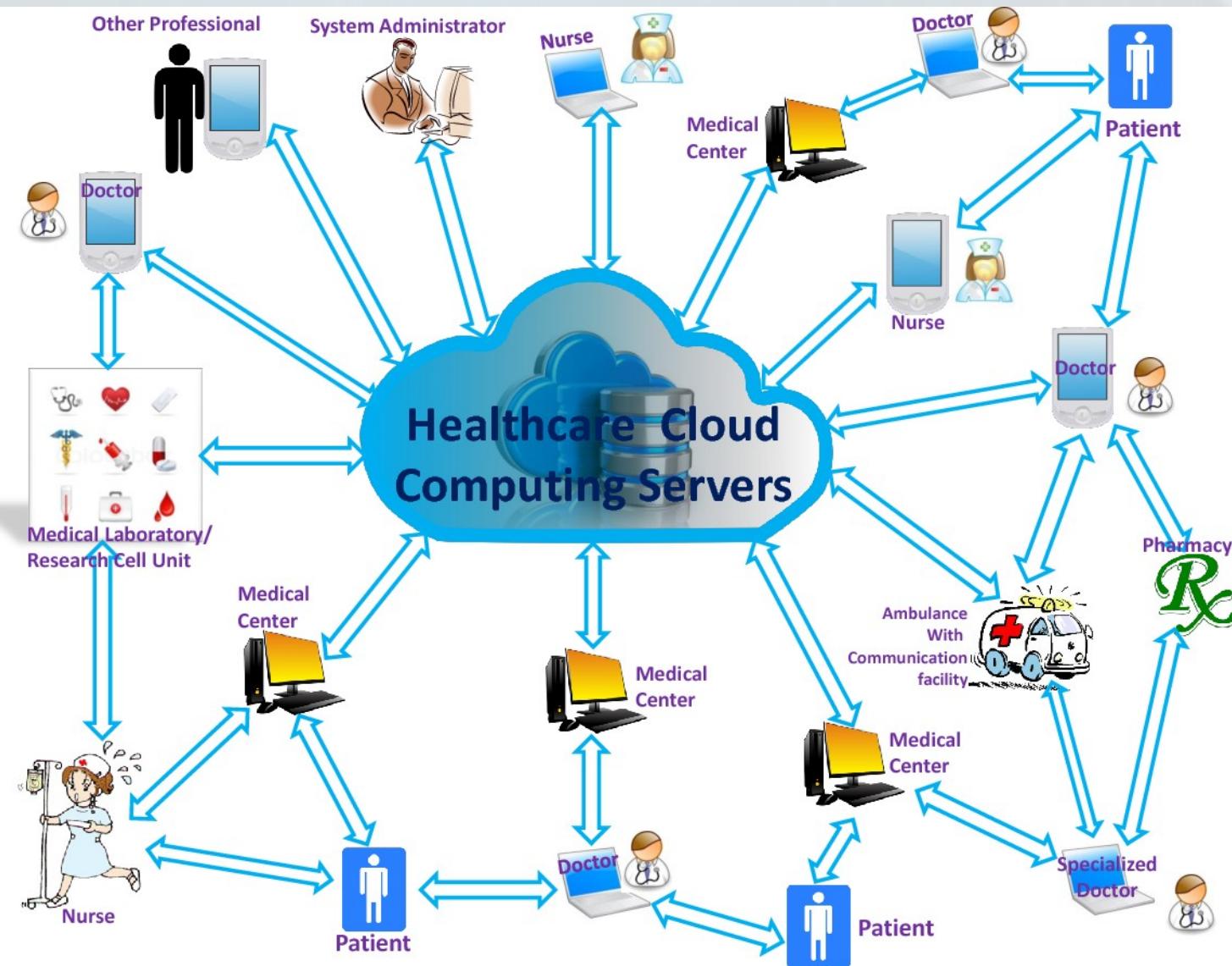
How it is related to medicine and healthcare system?

The top 10 medical advances in history

- Vaccines (1796)
- Anaesthesia (1846)
- Germ theory (1861)
- Medical imaging (1895)
- Penicillin (1928)
- Organ transplants (1954)
- Antiviral drugs (1960s)
- Stem cell therapy (1970s)
- Immunotherapy (1970s)
- **Artificial intelligence (21st century)**

How BIG?

- According to the National Hospital Discharge Survey (2010): there were 35.1 million discharges with an average length of 4.8 days a stay”
- According to the National Hospital Ambulatory Medical Care Survey (2011): there were 125.7 million outpatient visits and 136.3 million emergency department visits.”
- A human body generates around 2000 GB of data every single day.



Clinical Intelligence (CI)

30%

of US
hospitals

use a clinical data
warehouse/mining
solution, according to
HIMSS Analytics.

Business Intelligence (BI)

33%

of healthcare
organizations use
BI tools

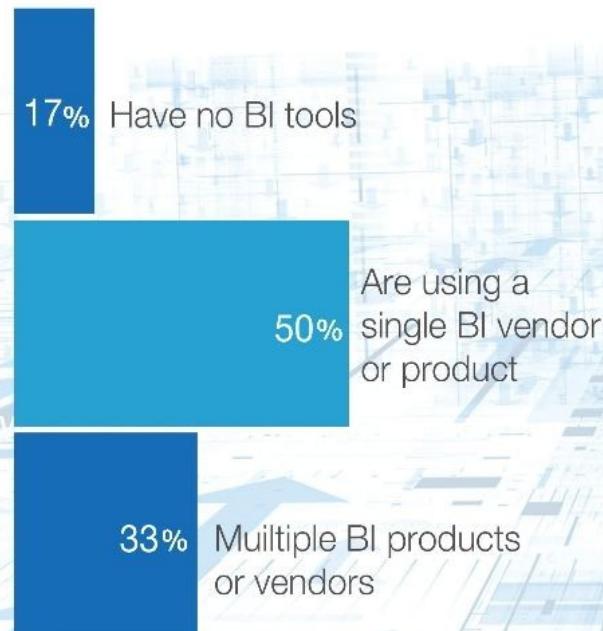
BENEFITS INCLUDE:

1. More cost-effective operations
2. Quality improvement.
3. Patient Satisfaction
4. Labor Costs

5 of the top purposes of clinical intelligence

- 1 Drive quality improvement programs
- 2 Identify individual care gaps
- 3 Stratify the population by level of risk
- 4 Measure long-term outcomes
- 5 Assess population health needs

How extensively are organizations using BI?



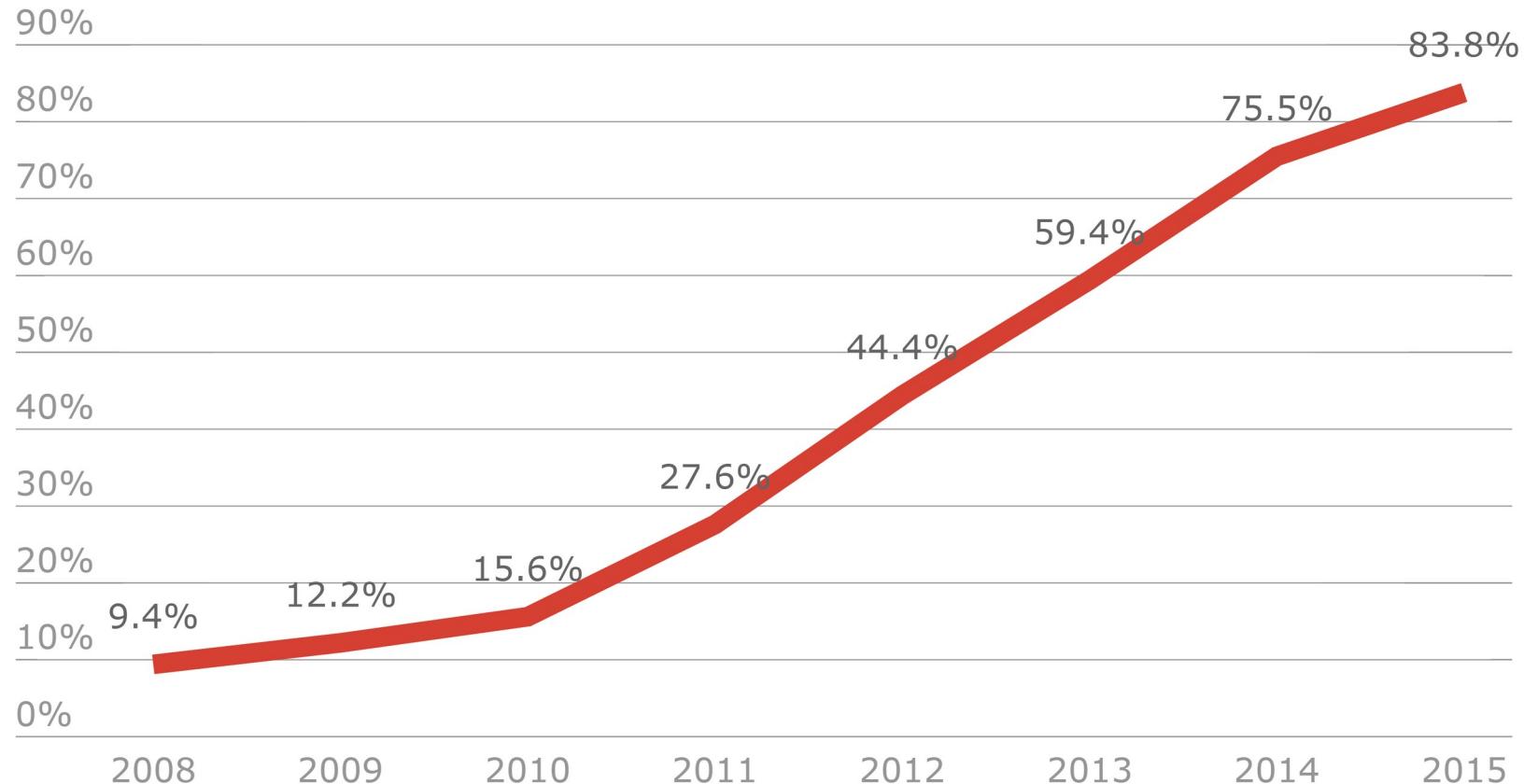
164,414

LIVES POTENTIALLY SAVED

A 2013 Healthgrades Hospital Clinical Excellence report found that, from 2009 through 2011, if all other hospitals performed at the level of Distinguished Hospitals for Clinical Excellence: 164,414 lives could have potentially been saved.

Rapid adoption of EHRs

% of acute-care hospitals with a basic electronic records system



Source: ONC/American Hospital Association

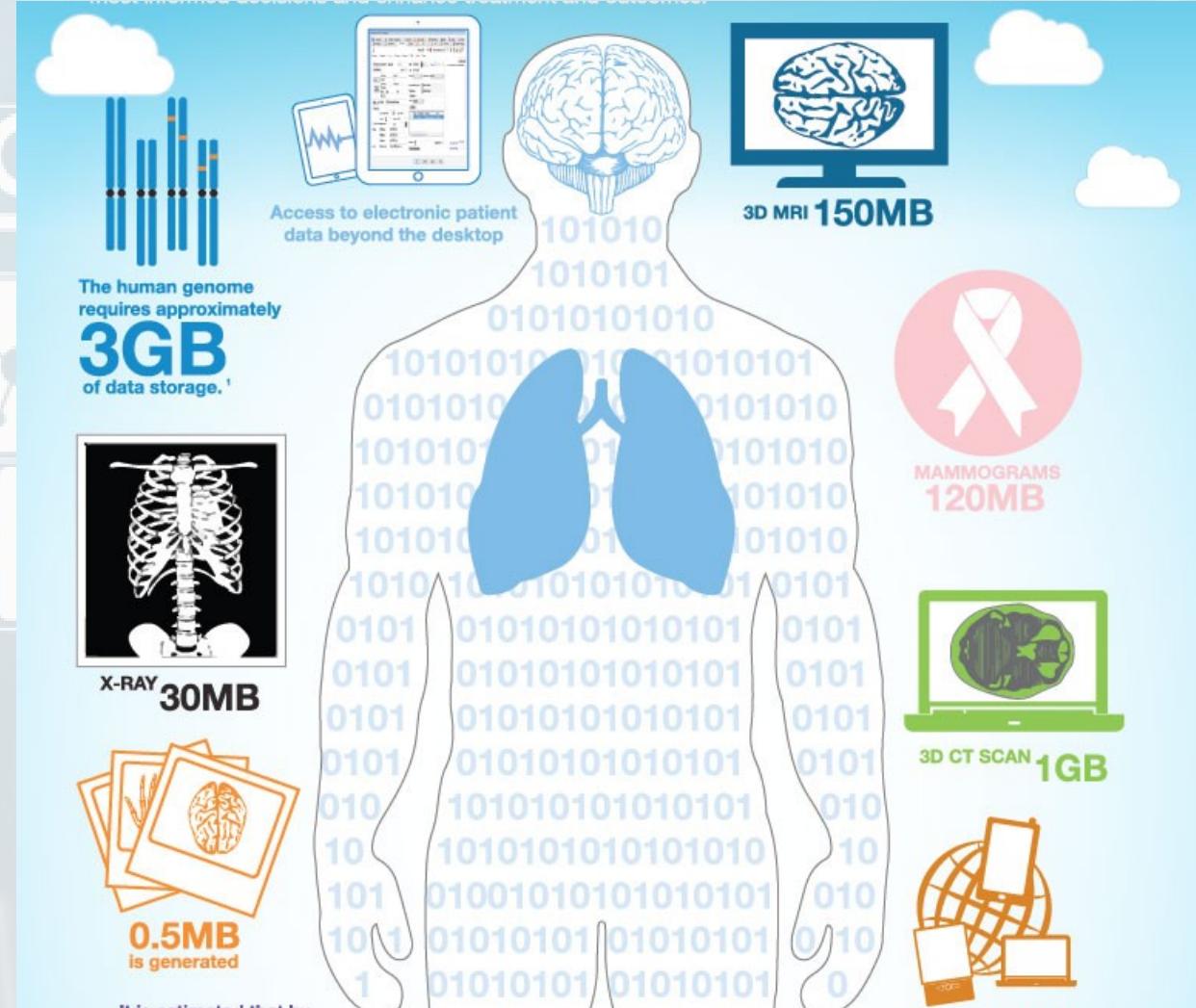
thanks
for
watching!

MEDICAL DATA SCIENCE

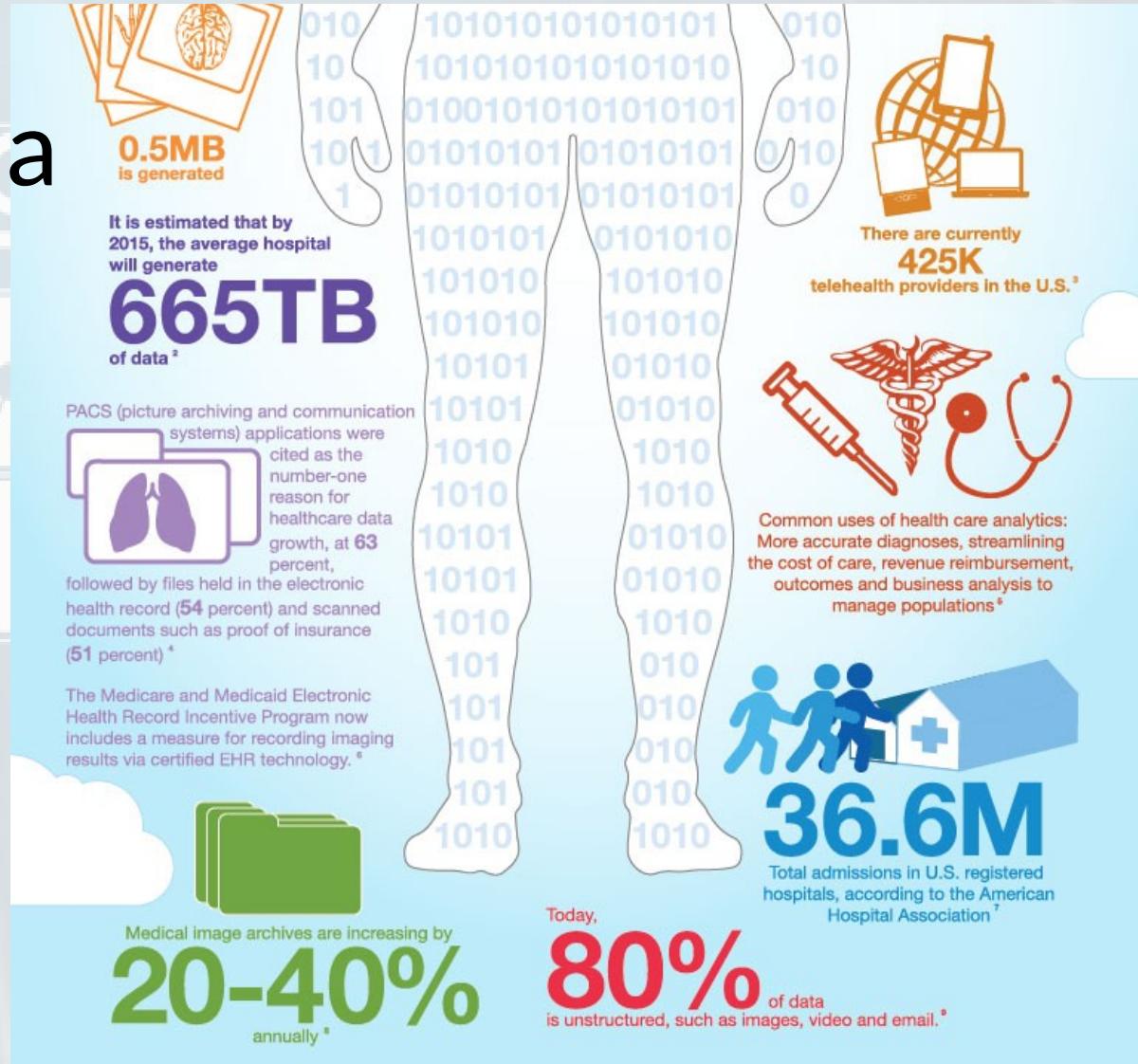
An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
9- Introduction:
E-health BIG data!

Our body is a source Of BIG data!



Our body is a source of BIG data!



Growth in healthcare data

1 exabyte = 1 billion gigabytes

2013
153
EXABYTES

Source: Stanford Medicine 2013, IDC 2014



2020
2,314
EXABYTES

16,000

hospitals worldwide
collect data on patients



4.9 million

patients worldwide will
use remote monitoring
devices by 2016¹



An 18%

annual compound
growth rate is anticipated
between 2010 and 2016
for patients that will use
remote monitoring
devices²

TO PUT THAT INTO PERSPECTIVE, DATA CENTERS GLOBALLY

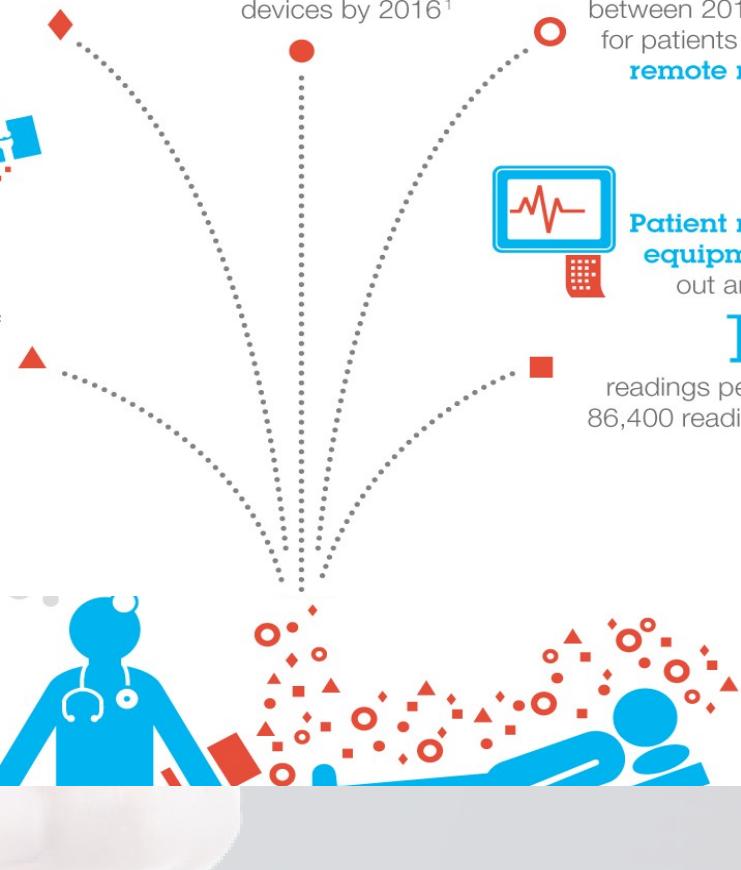
will only have enough room for an estimated

985 exabytes by 2020

meaning that almost two and a half times this capacity would be required to house all the healthcare data.

80%

of health data is unstructured and stored in hundreds of forms such as labs results, images, and medical transcripts



Patient monitoring equipment pumps out an average of

1,000

readings per second or
86,400 readings in a day

Big Data in Health Care

Generates Exabytes of Data

153
Exabytes in 2013

2,314
Exabytes in 2020

Source: "Harnessing the Power of Data in Health."

Genomic Sequencing

Medical Imaging

Health Records

Payor Records

Medical Devices

Smartphones

Wearables

Search Engine Data

Pharmaceutical Research

+

+

+

+

+

+

+

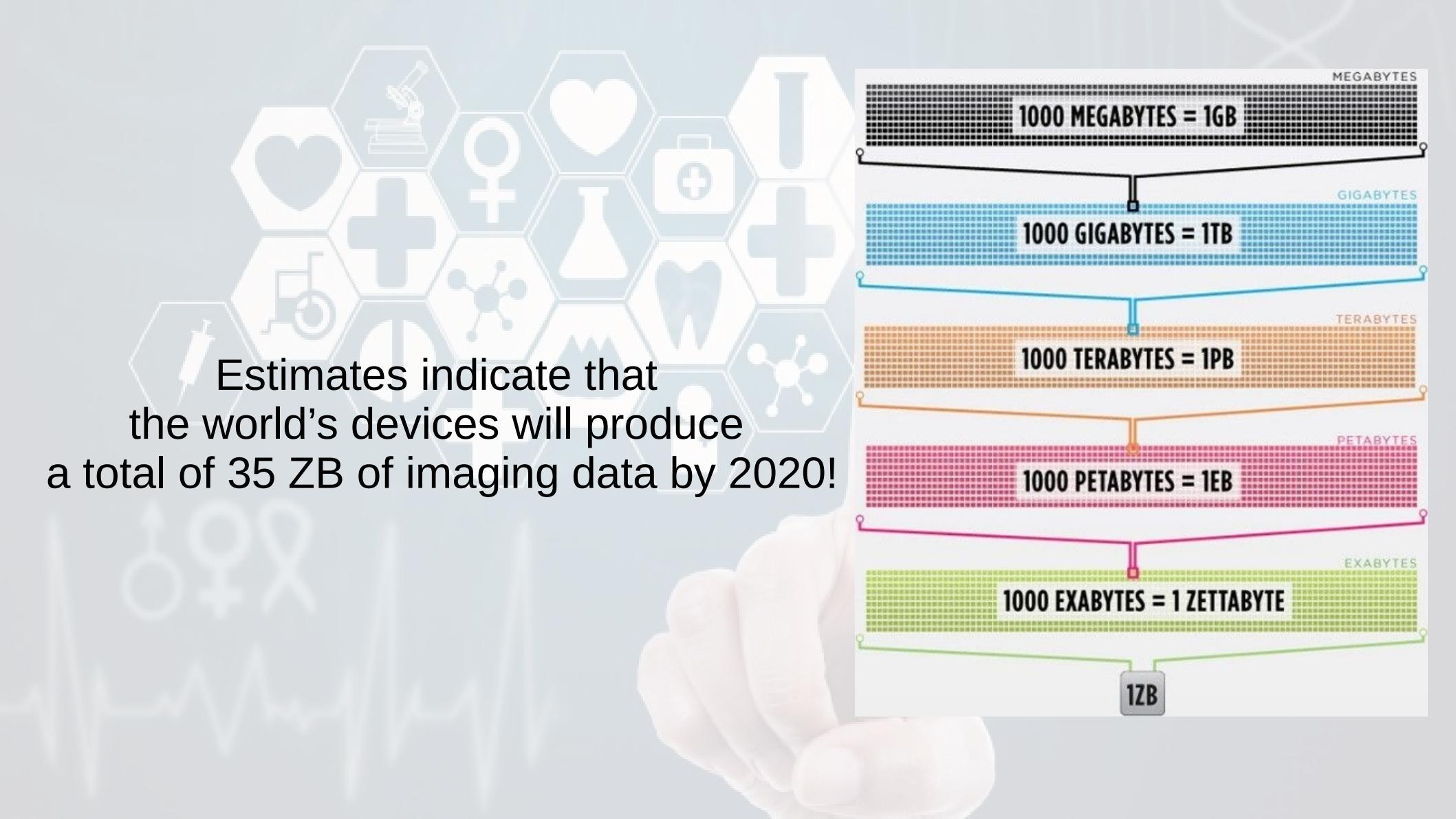
+

+

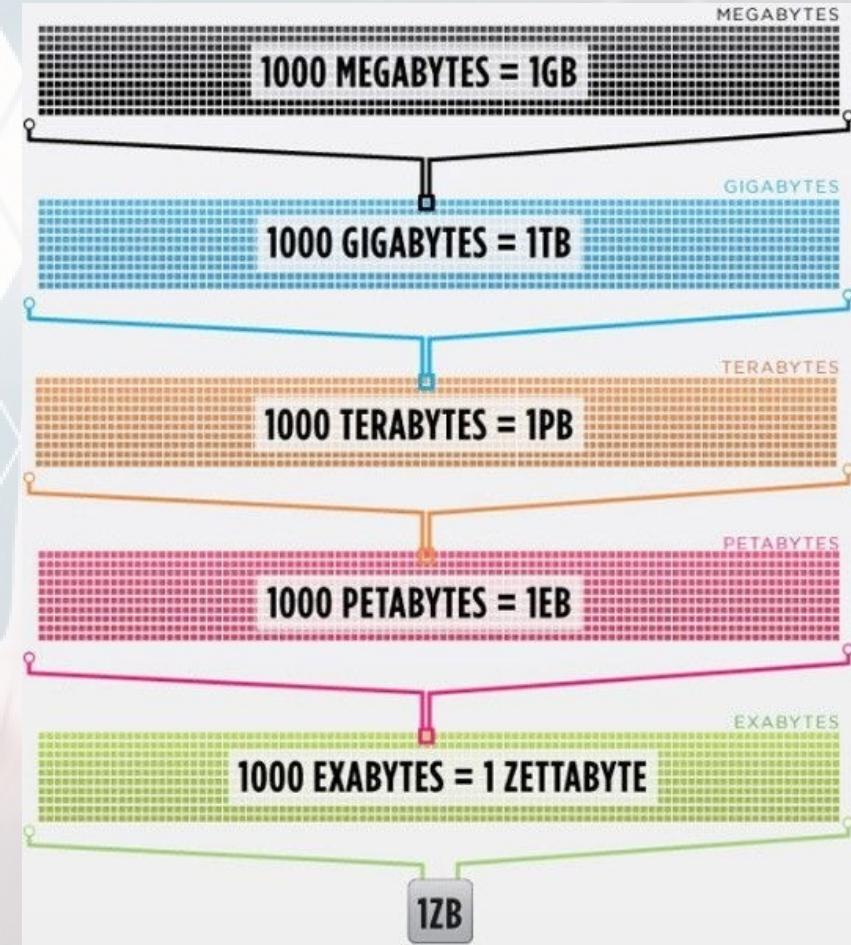
+

+

+



Estimates indicate that
the world's devices will produce
a total of 35 ZB of imaging data by 2020!





“According to the National Academics of sciences and engineering, about 12 million adult patients are misdiagnosed each year in the US.”



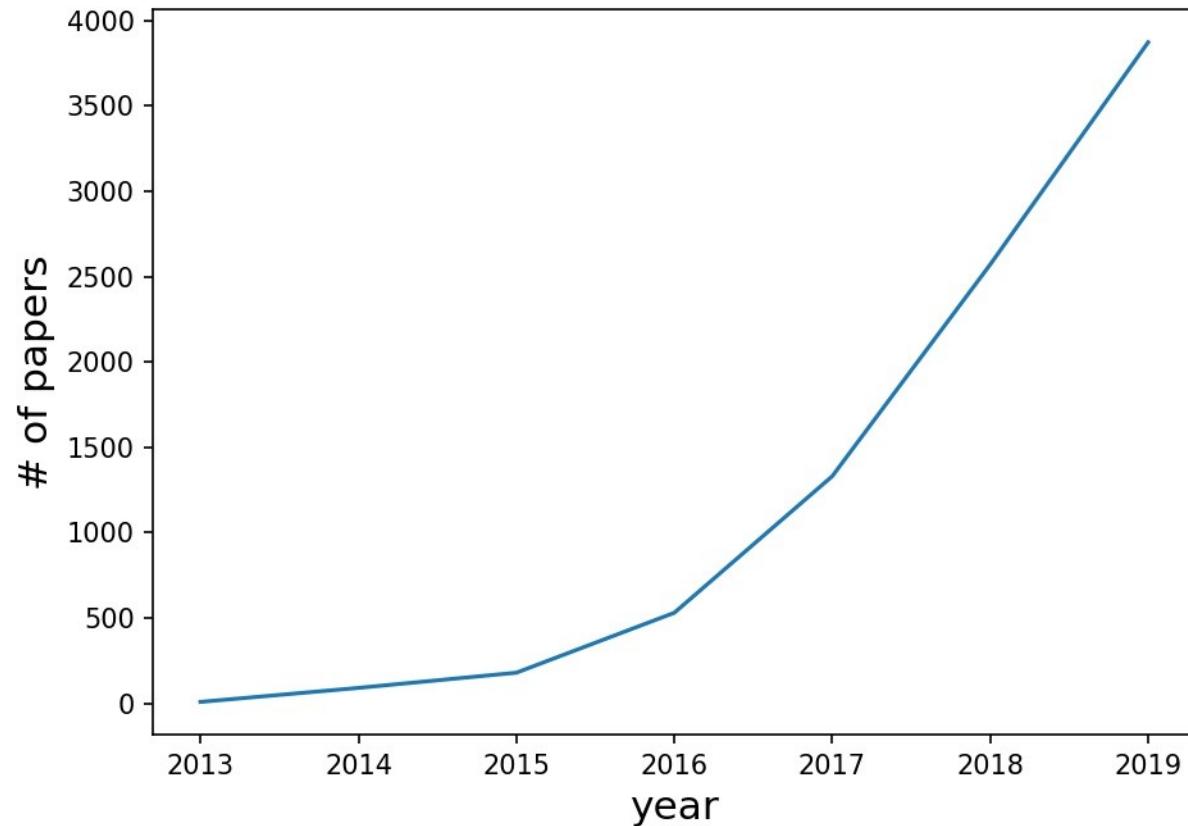
thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
10- Introduction:
It is getting hotter!

Number of papers that include “Machine learning” in the manuscript (biorxiv.org)





Over the past 5 years
AI base health companies received
a total of more than \$1.8 billion investments
in 446 separate transactions.

- Hot Market Trend 1: AI and ML
- Hot Market Trend 2: Blockchain
- Other Segments Still Gaining Momentum

Distributed and secure technologies

2 Blockchain (10%)

Includes companies that design and develop health data aggregators and big data analysis

3 Cloud-based (8%)

Includes companies that design and develop clinical data exchanges within an enterprise, image repositories, and enterprise viewers

Real-time and improved imaging

4 3D visualization and VR (13%)

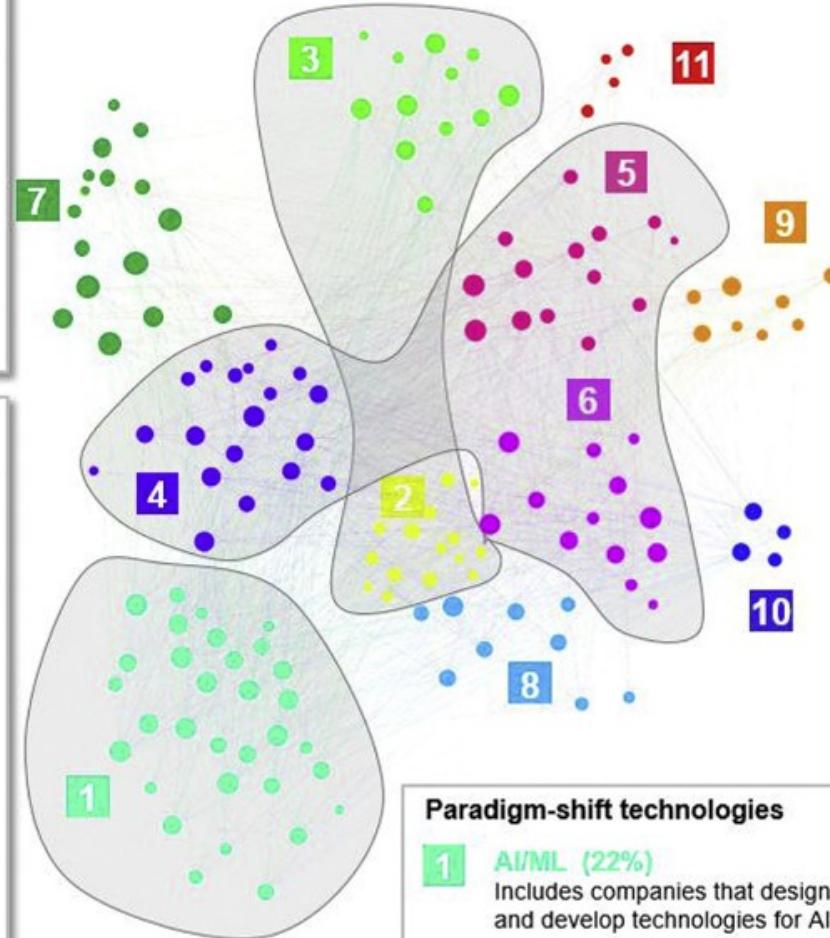
Includes companies that design and develop ultrasound technologies and medical imaging

5 Intraoperative (10%)

Includes companies that design and develop equipment for precision surgical guidance, MRI guidance, and wire-guided catheters

6 Nuclear and hybrid imaging (9%)

Includes companies that design and develop nuclear image technology and photosensitizing compounds to help identify cancer in medical images



Paradigm-shift technologies

1 AI/ML (22%)

Includes companies that design and develop technologies for AI and deep-learning algorithms that interpret images

Incremental innovation and point solutions

7 Portable (11%)

Includes companies that design and develop portable ultrasound technologies and medical imaging

8 Faster acquisition (6%)

Includes companies that design and develop ultrasonic non-invasive sensors and micro ultrasound machine for prostate biopsies

9 Clinical decision support (5%)

Includes companies that design and develop image analysis to help doctors make decisions

10 Telemedicine (3%)

Includes companies that enable doctors and patients to interact remotely

11 3D printing (3%)

Includes companies that design and develop personalized 3D models and medical image segmentation

	Company count	Average number of transaction per company ¹	Median transaction value (\$ million)	Total transaction value (\$ million)	Number of companies generating revenue	Average revenue (\$ million)	Number of profitable companies	Average year founded
Paradigm shift	1 AI/ML	32	2.9	6.8	555.8	12	30-50	1 2011
Distributed and secure technologies	2 Blockchain	15	1.5	0.5	71	6	<5	0 2013
	3 Cloud-based	12	3.2	3.5	194.4	6	<5	0 2007
Real-time and improved imaging	4 3D visualization and VR	19	3.7	7.4	197	9	15-30	1 2006
	5 Intraoperative	14	4.6	3.3	364.9	5	5-15	1 2010
	6 Nuclear and hybrid imaging	13	4.1	3.8	142.1	5	<5	1 2010
Incremental innovation and point solutions	7 Portable	16	2	5.3	176.8	5	<5	0 2006
	8 Faster acquisition	9	3.4	5.5	57.4	5	<5	0 2007
	9 Clinical decision support	8	3.9	3.8	54.6	5	<5	1 2005
	10 Telemedicine	4	1.8	1.3	2.7	4	<5	0 2010
	11 3D printing	4	1	25	25	1	<5	0 2012

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in healthcare

By: Alireza Vafaei Sadr
11- Introduction:

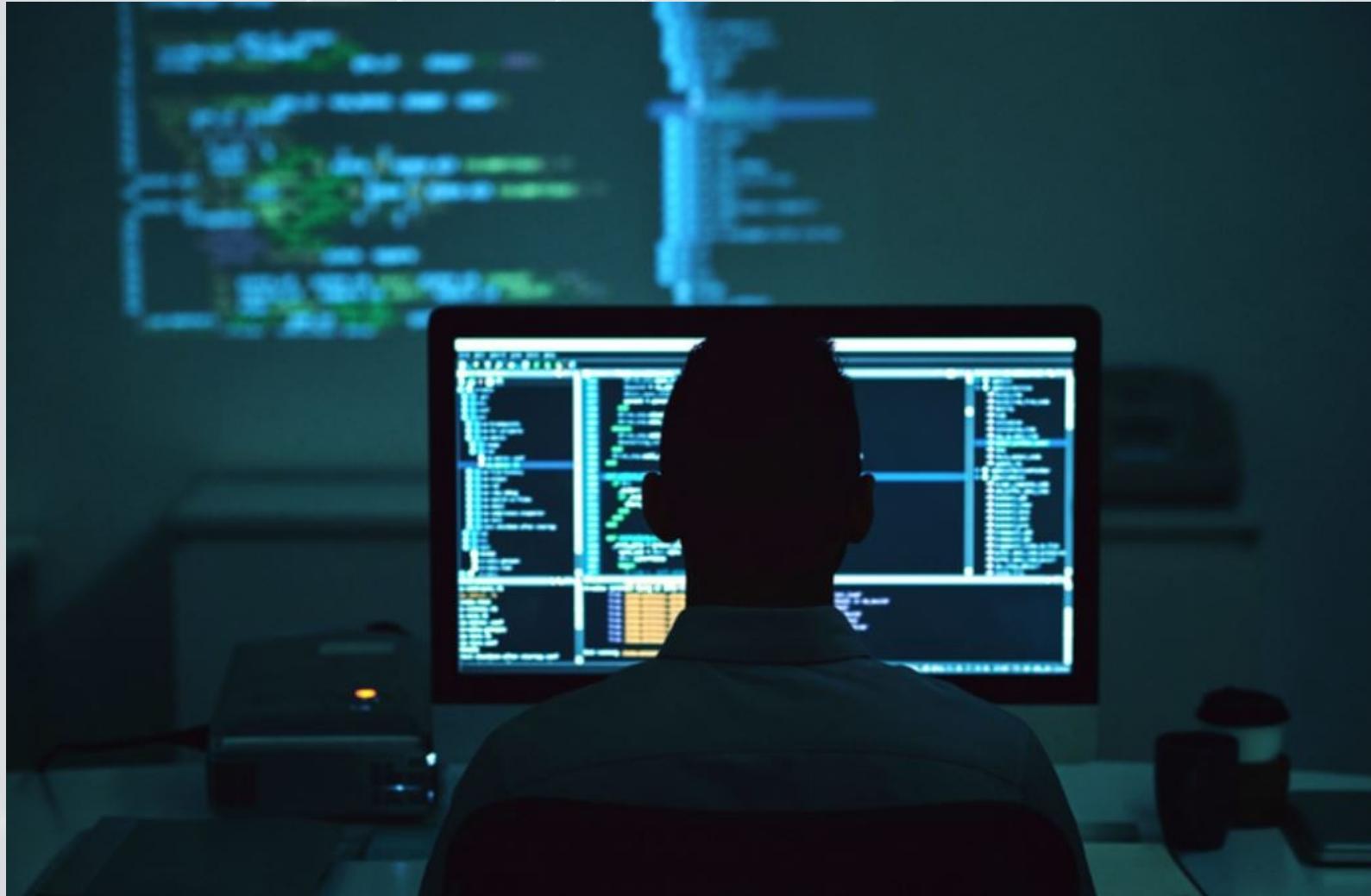
What do you need to begin data science?



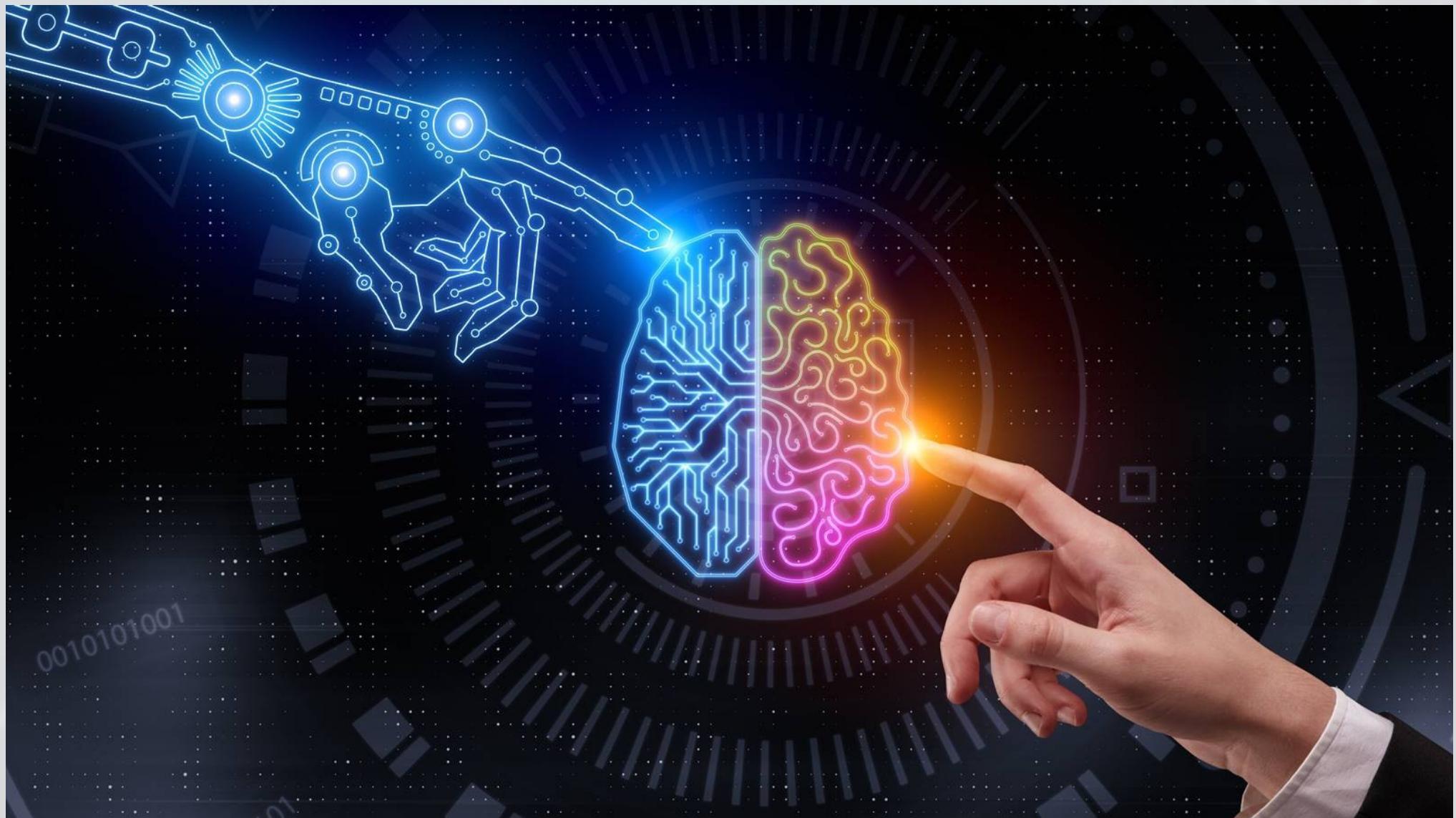
**Data
Programming
Statistics
AI**

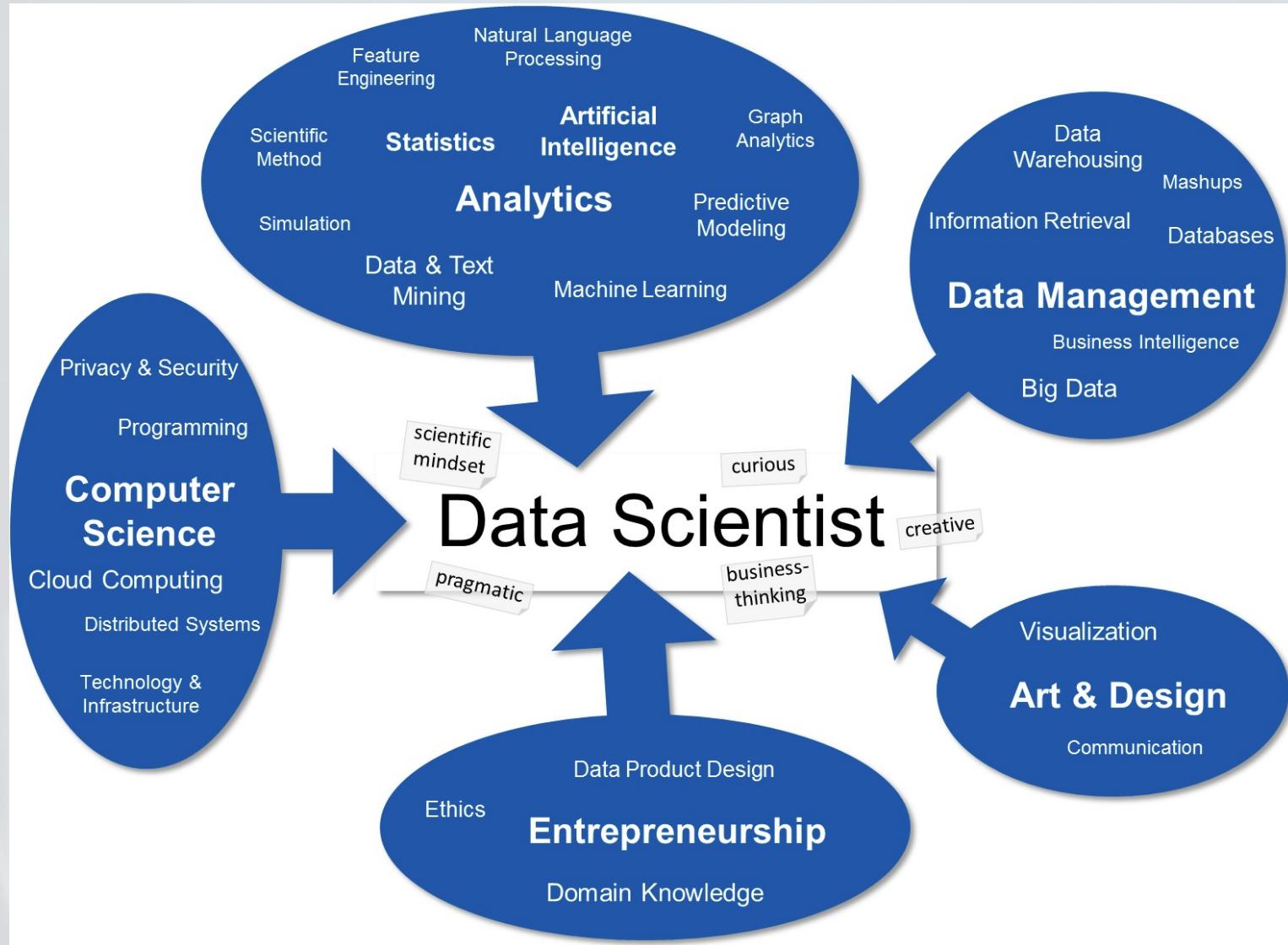
And some Meta-skills











You definitely need to
(Computer+“book”)
Be a ~~Book~~ Worm!



MakeAGIF.com

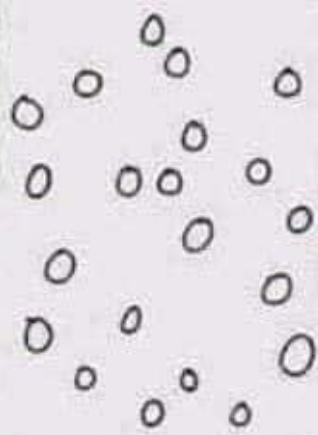
thanks
for
watching!

MEDICAL DATA SCIENCE

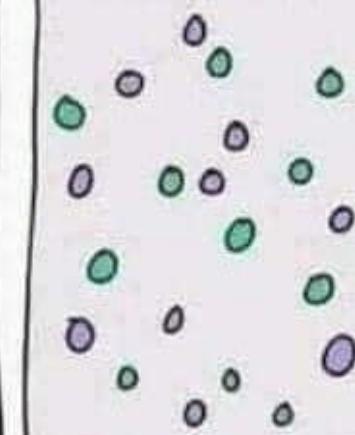
An introduction to:
data science applications in Search Healthcare

By: Alireza Vafaei Sadr
12- Introduction:
Data!?

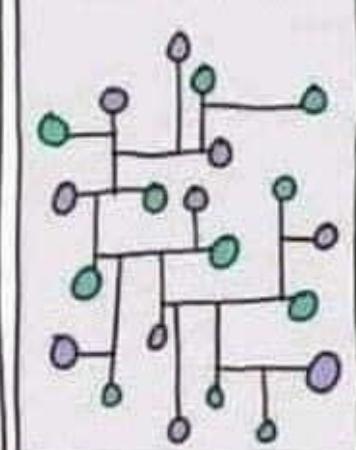
Data



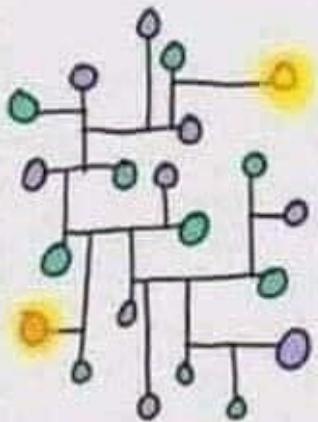
Information



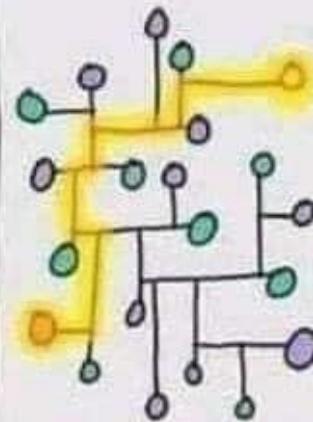
Knowledge



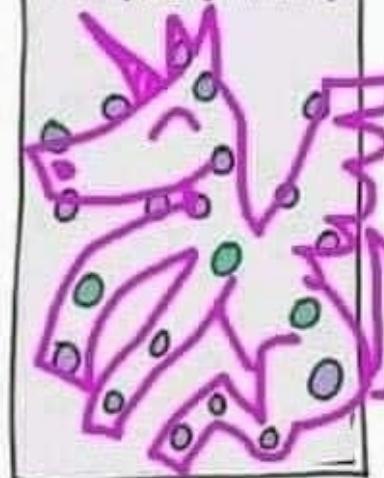
Insight



Wisdom



Conspiracy Theory



data acquisition:

- Data Sources: Companies/Proprietary Data, APIs, Government, Academic, Web Scraping/Crawling

Types of data

- Structured vs. Unstructured
- Quantitative vs. Categorical
- Discrete vs. Continuous
- Ordinal vs. Nominal

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in Search Healthcare

By: Alireza Vafaei Sadr
13- Introduction:
Why programming?

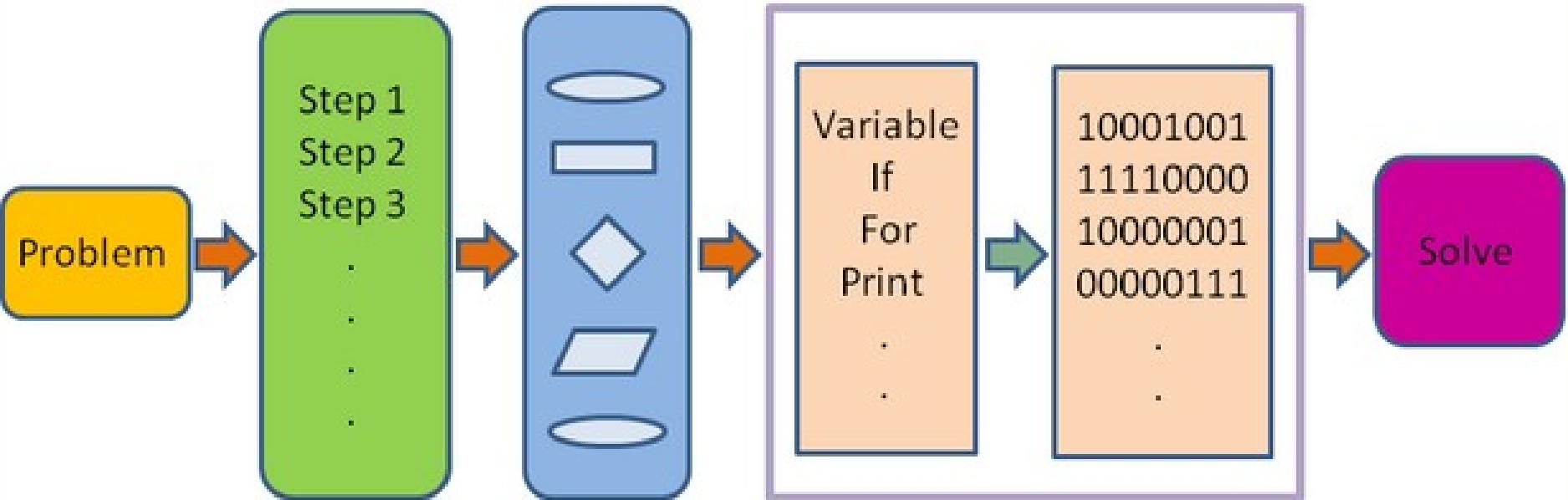
Why is programming important?

It's a simple question. There's a simple answer.

- We are entering an era of history where everything we do is assisted by (or simply done by) technology.
- The workplace will be filled with technology. It would be logical if we taught children how to use it and program it.
- Knowing how to program is important to every student if they want a greater chance of succeeding in the future.
- Students are falling behind in technological knowledge, and potential careers are blocked by lack of computer education at a young age. You will never know if you like something if you never get to try it.

Language?



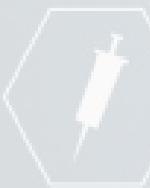


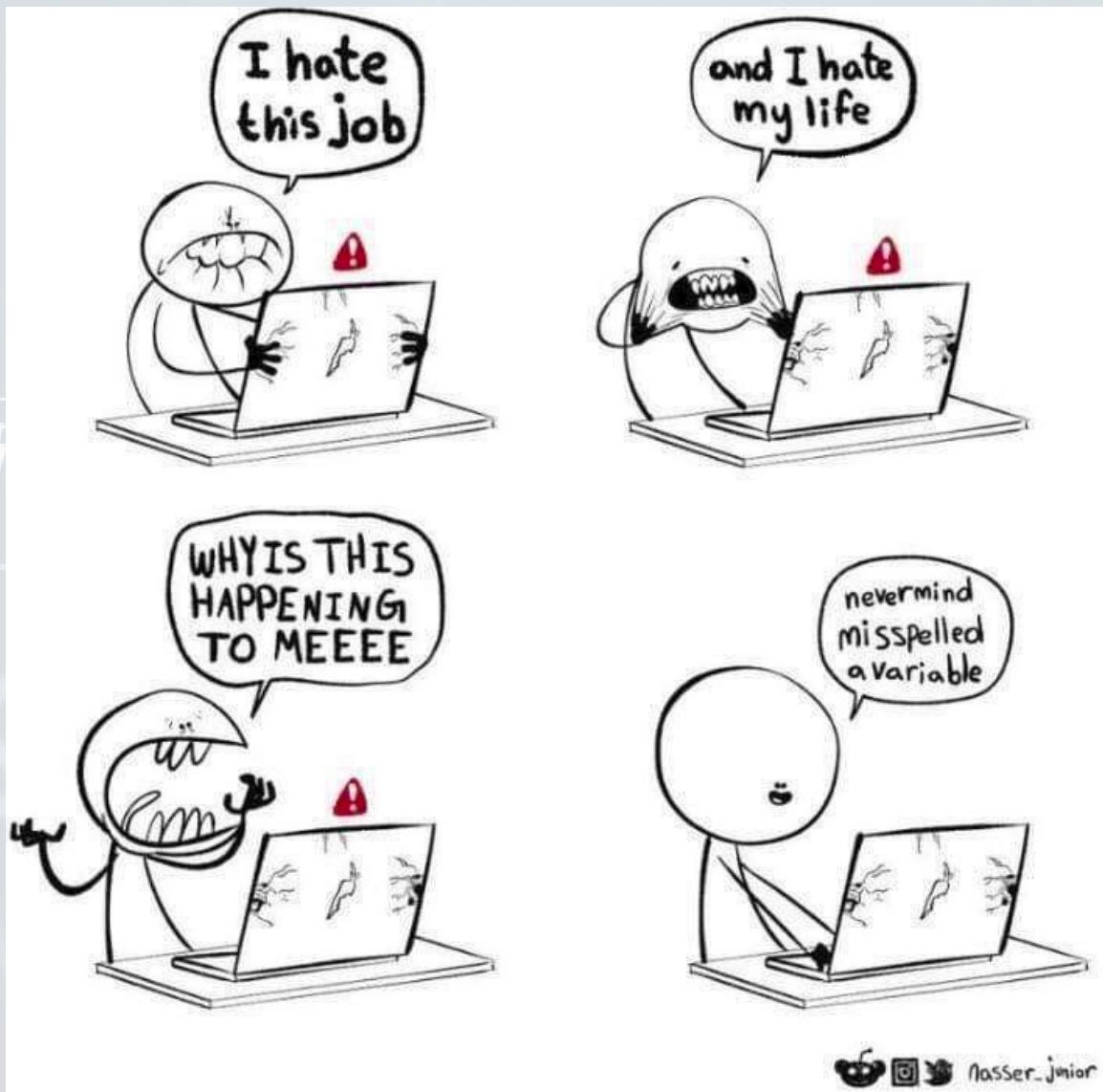
Algorithm

Flowchart

Programming language

(Basic , VB , C , C++ , C# , Java , Perl , ...)





thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in Search Healthcare

By: Alireza Vafaei Sadr
14- Introduction:
Data science language?

Let's ask **Google**: "the best programming language for data science"





What is in the beginner course?

- An Example!
- What is Python?
- How do we install python?
- Dive into programming (Don't worry!)



- (It is similar to our language!) High level
- Best language for data science
- Free
- It needs less coding
- It's able to talk to the others!
- Interactive, Interpreted, Modular, Portable, Dynamic, Object-oriented

7 Reasons Why You Should Learn Python



01

Perfect For Rookies

02

Community

03

Career Opportunities

04

Python in Web Development

05

Python in Artificial Intelligence and Machine Learning

06

Raspberry Pi

07

Startups and Corporates-
Python for Both

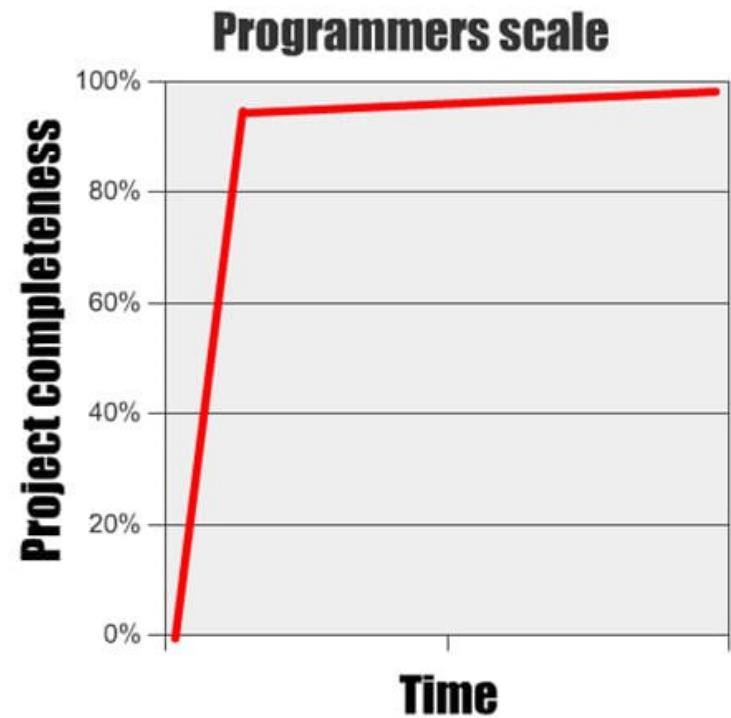


Guido Van Rossum



python SOFTWARE FOUNDATION

The 15 Stages of programming:	Happiness (%)
The beginning	50
I have no Idea what I'm doing	20
I should have been an English major	10
This makes more sense now	45
Almost ready to debug	60
Clean Compile!	95
Segmentation Fault	NULL
debugging	5
Cheap [REDACTED]	70
I should have finished this last night	5
Almost done	80
Finished!	100
Forgot to add that little feature	35
Seg Fault!	NULL
this I'm playing Skyrim	100





**NOT SURE IF I AM A GOOD
PROGRAMMING**

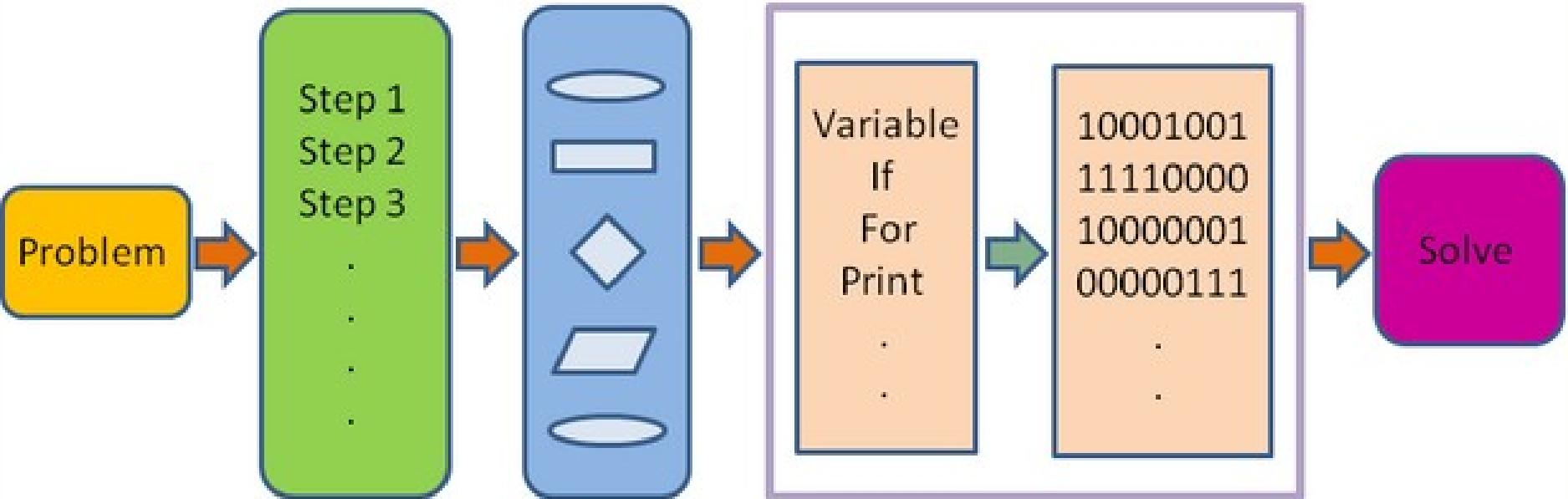
**OR GOOD AT
GOOGLING**

thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in Search Healthcare

By: Alireza Vafaei Sadr
16- Python:
How does Python work?

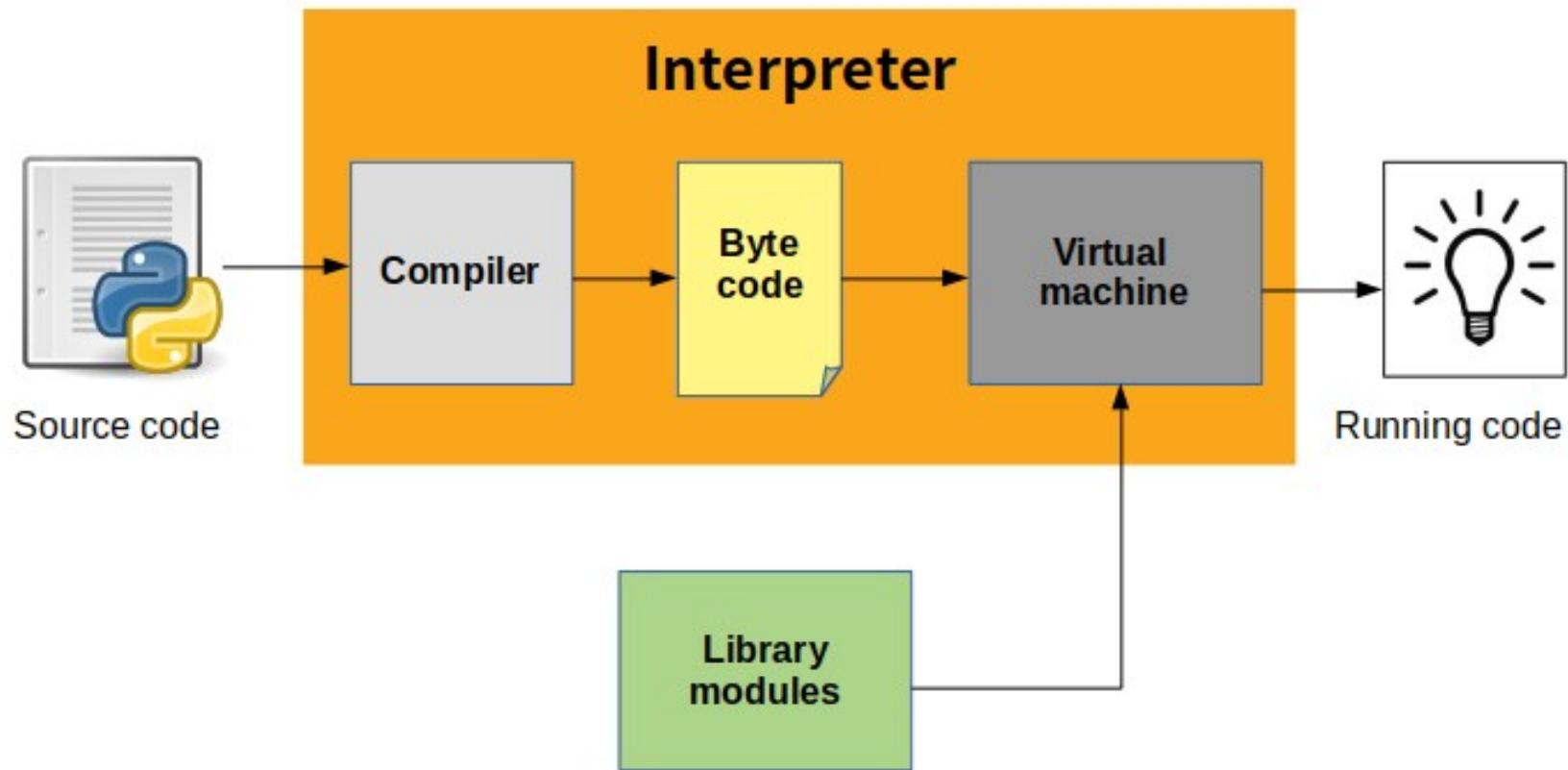


Algorithm

Flowchart

Programming language

(Basic , VB , C , C++ , C# , Java , Perl , ...)



thanks
for
watching!

MEDICAL DATA SCIENCE

An introduction to:
data science applications in Search Healthcare

By: Alireza Vafaei Sadr
17- Python:
Anaconda



CONDA[®]



MINICONDA[®]

= conda
+ python
+ base packages



ANACONDA[®]

= miniconda
+ 150 high quality packages

thanks
for
watching!