



INDIAN INSTITUTE OF
INFORMATION
TECHNOLOGY

Data Science Systems - Motivation

Dr. Animesh Chaturvedi

Assistant Professor: IIIT Dharwad

Post Doctorate: King's College London & The Alan Turing Institute

PhD: IIT Indore MTech: IIITDM Jabalpur



Indian Institute of Technology Indore
भारतीय प्रौद्योगिकी संस्थान इंदौर

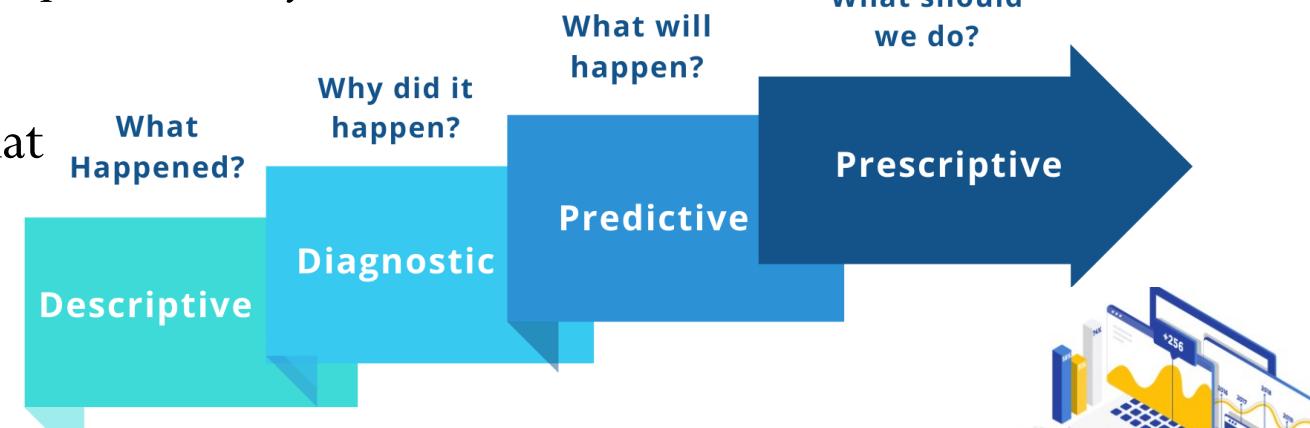


PDPM
Indian Institute of Information Technology,
Design and Manufacturing, Jabalpur

The
Alan Turing
Institute

Data Science and Data Analytics

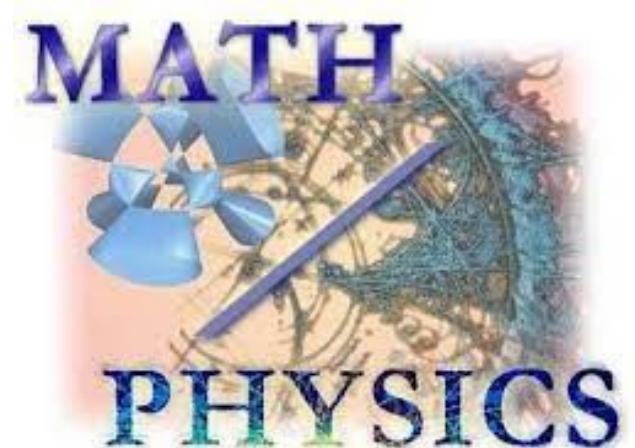
- **Data Science:** interdisciplinary science that
 - deals with data: methods, algorithms, processes, systems etc.
 - Theory oriented
- **Data Analytics:** analysis of data that
 - discovers trends, graph, tables etc.
 - Technology oriented
- Both
 - extracts knowledge and apply actions through structured and unstructured data insights
 - deals with data mining, machine learning, data management, and big data.
- **Data Scientist** and **Data Analyst** applies Data Science and Data Analytics



Data Science as Interdisciplinary Science

Discipline Examples

- **Statistical physics:** methods of probability theory and statistics, and mathematical tools
 - deals with large populations and approximations,
- **Information theory:** quantification, storage, and communication of information (e.g. signals)
 - Subfield of Communication, Electronics, and Computer Science
- **Nonlinear dynamics:** the change of the output is not proportional to the change of the input.
 - Changes in variables over time, or, space etc.
 - Measures chaotic, unpredictable, or counterintuitive



Discipline Examples

- **Computer science:** programming creates complex systems, and application involving complex real-world data,
- **Sociology:** studies human behaviour, social behavior, society, patterns of social relationships, social interaction, and culture
- **Social network:** investigate social structures with networks and graph theory
- **Biology:** studies life and living organisms, physical structures of chemical, molecular interactions
- **Data Science:** interdisciplinary field that uses statistics, computing, scientific methods, processes, algorithms, social sciences, and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data.

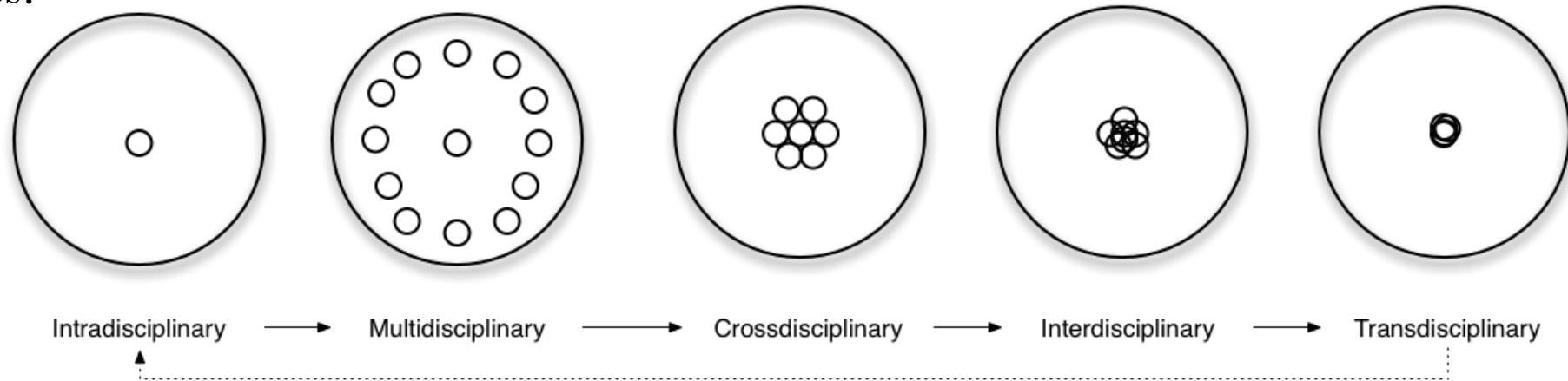
Interdisciplinary

- An organizational unit involving two or more academic disciplines,
- Dedicated journals, conferences and university departments.
- Three levels of cross-disciplinary research:
 - **Multidisciplinarity:** Pluridisciplinary level draws knowledge from different disciplines but stays within their boundaries.
 - **Interdisciplinarity:** Cross-disciplinary level analyzes, synthesizes and harmonizes links between disciplines.
 - **Transdisciplinarity:** Discipline-forming level integrates and transcends traditional boundaries.

[8] Choi BC, Pak AW. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. Clin Invest Med. 2006 Dec; 29(6):351-64. PMID: 17330451.
<https://en.wikipedia.org/wiki/Interdiscipline>

Disciplinaries: intra, cross, multi, inter, trans

- Intradisciplinary: working within a single discipline.
- Crossdisciplinary: viewing one discipline from the perspective of another.
- Multidisciplinary: people from different disciplines working together, each drawing on their disciplinary knowledge.
- Interdisciplinary: integrating knowledge and methods from different disciplines, using a real synthesis of approaches.
- Transdisciplinary: creating a unity of intellectual frameworks beyond the disciplinary perspectives.

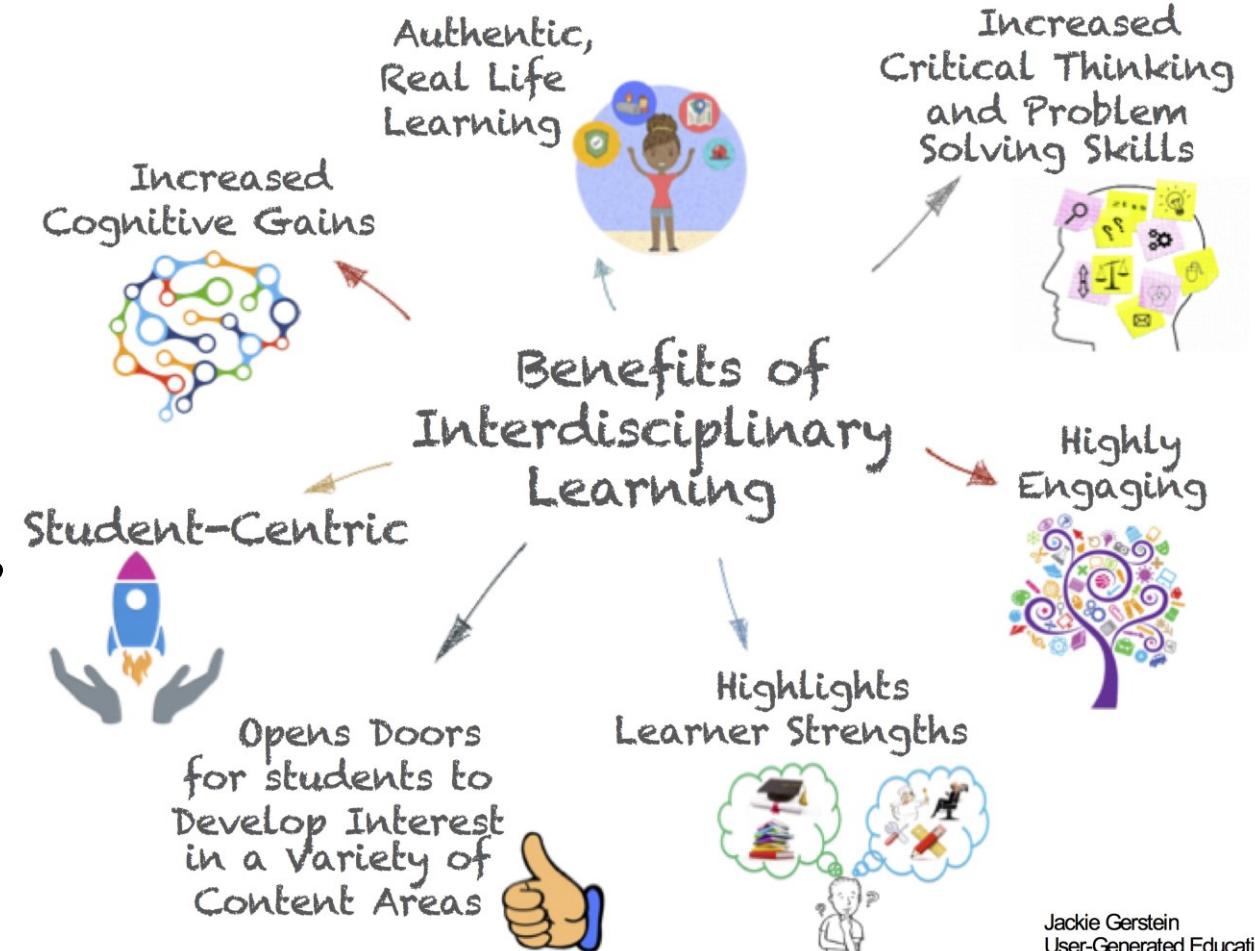


Interdisciplinary

- aka. Interdisciplinary studies
- combination of two or more academic disciplines into one activity (e.g., a research project).
- Disciplines could be like
 - social science,
 - mechanical engineering,
 - electrical engineering,
 - computer science and engineering, etc.
- Inter-discipline: Electromechanics, Mechatronics, Bioinformatics, Biomedical Engineering, Computational Social Systems etc.
- **Such disciplines produce data for analysis**

Interdisciplinary Examples

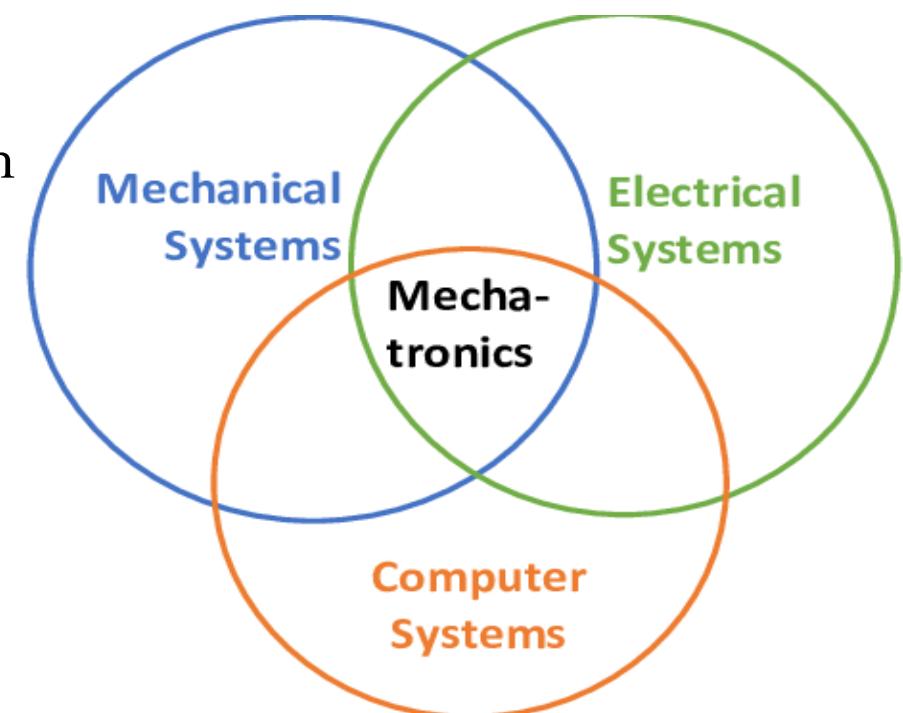
- Electromechanics focuses on the interaction of electrical and mechanical systems as a whole and how the two systems interact with each other.
- Mechatronics focuses on the engineering of electronic, electrical and mechanical engineering systems, and also includes a combination of robotics, electronics, computer, telecommunications, systems, control, and product engineering.



<https://en.wikipedia.org/wiki/Electromechanics>
<https://en.wikipedia.org/wiki/Mechatronics>

Electromechanics and Mechatronics

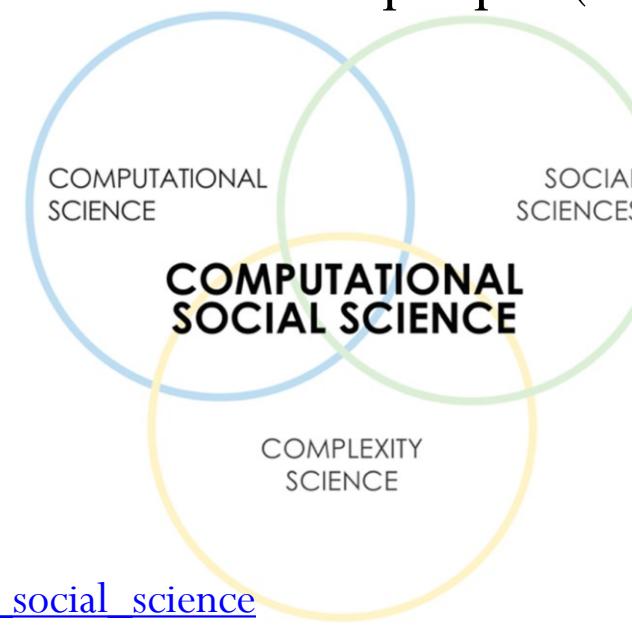
- An actuator is a component of a machine that is responsible for moving and controlling a mechanism or system, for example by opening a valve. In simple terms, it is a "mover".
- An actuator requires a control signal and a source of energy.
- A sensor is a device, module, machine, or subsystem whose purpose is to detect events or changes in its environment and send the information to other electronics, frequently a computer processor. A sensor is always used with other electronics.



<https://en.wikipedia.org/wiki/Electromechanics>
<https://en.wikipedia.org/wiki/Mechatronics>
<https://en.wikipedia.org/wiki/Actuator>

Interdisciplinary Examples

- Computational social science is sub-disciplines concerned with computational approaches to the social sciences.
 - Sub-field, Natural Language processing
- Human–Computer Interaction (HCI) studies the design and use of computer technology, focused on the interfaces between people (users) and computers.
 - SixthSense

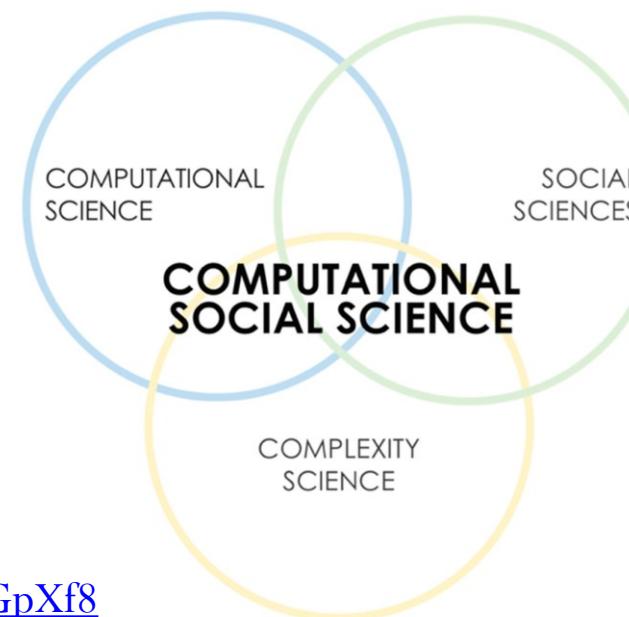


https://en.wikipedia.org/wiki/Computational_social_science

https://en.wikipedia.org/wiki/Human-computer_interaction

Sixth Sense

- Gesture-based wearable computer system developed at MIT Media Lab by
 - Steve Mann in 1994 and 1997 (headworn gestural interface),
 - 1998 (neckworn version), and
 - Pranav Mistry (also at MIT Media Lab), in 2009, developed both hardware and software for headworn and neckworn versions of it.

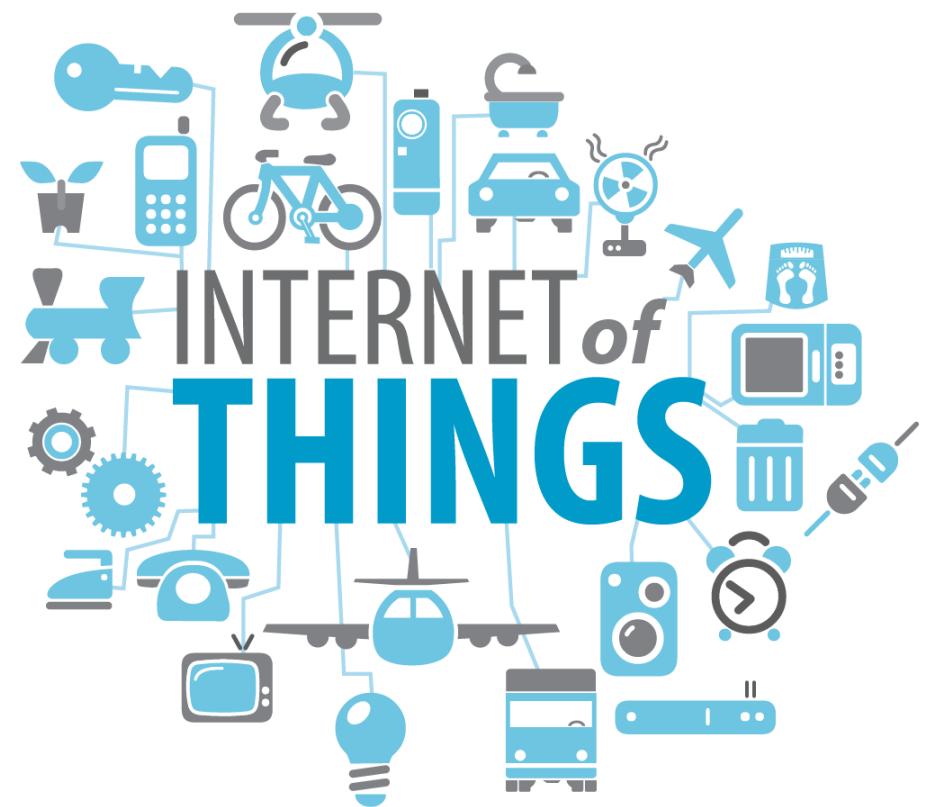


<https://www.youtube.com/watch?v=E8GU-dGpXf8>

<https://en.wikipedia.org/wiki/SixthSense>

Internet of things (IoT)

- The Internet of things (IoT) is the inter-networking of physical devices, embedded with **electronics, software, sensors, actuators, and network connectivity** which enable these objects to collect and exchange data.



Natural Language Processing

- Text data mining or analytics is the process of analyzing quality information from text; by automatically discovering and extracting unknown information from text resources.
 - Regular Expressions (Regex), Text Pattern matching on Textual Data or (NLP)
- Sentiment analysis involve analysis of labeled or unlabeled natural language based affectivity of words and concepts made from WordNet and ConceptNet.
 - Positive, Negative, and Neutral Sentiments

Case Study: Bioinformatics Data

Sub-branches of BME

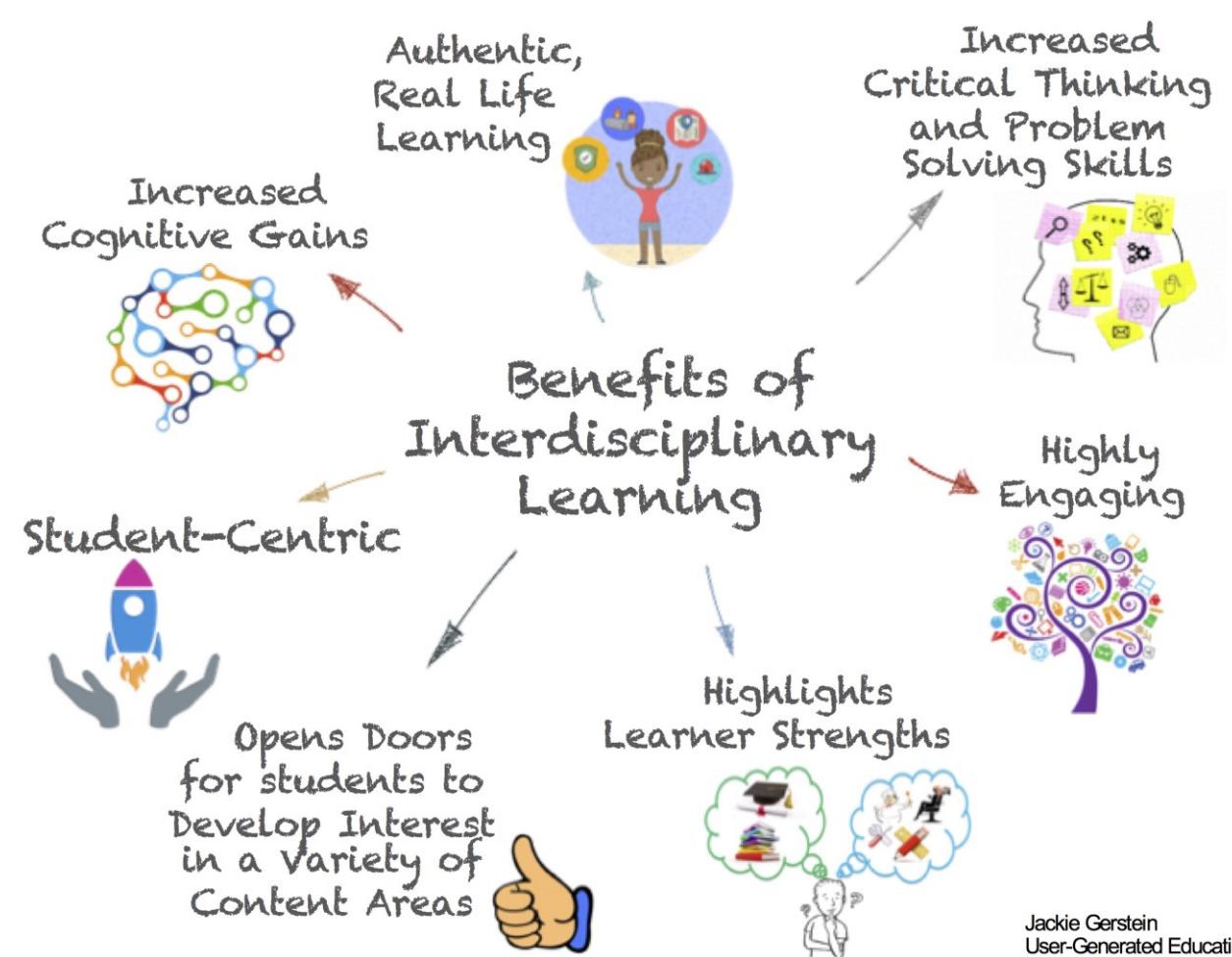
- **Bioinformatics (Case study for Data analysis)**
- Biomechanics
- Biomaterials science or engineering
- Biomedical optics
- Tissue engineering
- Genetic engineering
- Neural engineering
- Pharmaceutical engineering
- Medical devices (Medical imaging, Implants, Bionics, and Biomedical sensors)
- Clinical engineering
- Rehabilitation engineering

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

<https://en.wikipedia.org/wiki/Bioinformatics>

Biomedical engineering and Bioinformatics

- Biomedical engineering (BME) is the application of engineering principles and design concepts to medicine and biology for healthcare purposes (e.g., diagnostic or therapeutic).
- Bioinformatics develops methods and software tools for understanding biological data, in particular when the data sets are large and complex.



https://en.wikipedia.org/wiki/Biomedical_engineering
<https://en.wikipedia.org/wiki/Bioinformatics>

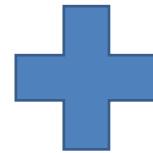
Bioinformatics data analysis

- **Informatics:** set of digital codes and a language
- **Bioinformatics:** Study of biological (or life) information (digital code for studying properties of bio-systems)

**Computer scientists,
Mathematicians, Data
Scientist etc.**

Develop tools, software,
algorithms

Store and analyze the data.



Biologists

collect molecular data:
DNA & Protein
sequences,
gene expression, etc.



Bioinformaticians

Study biological
questions by analyzing
molecular data

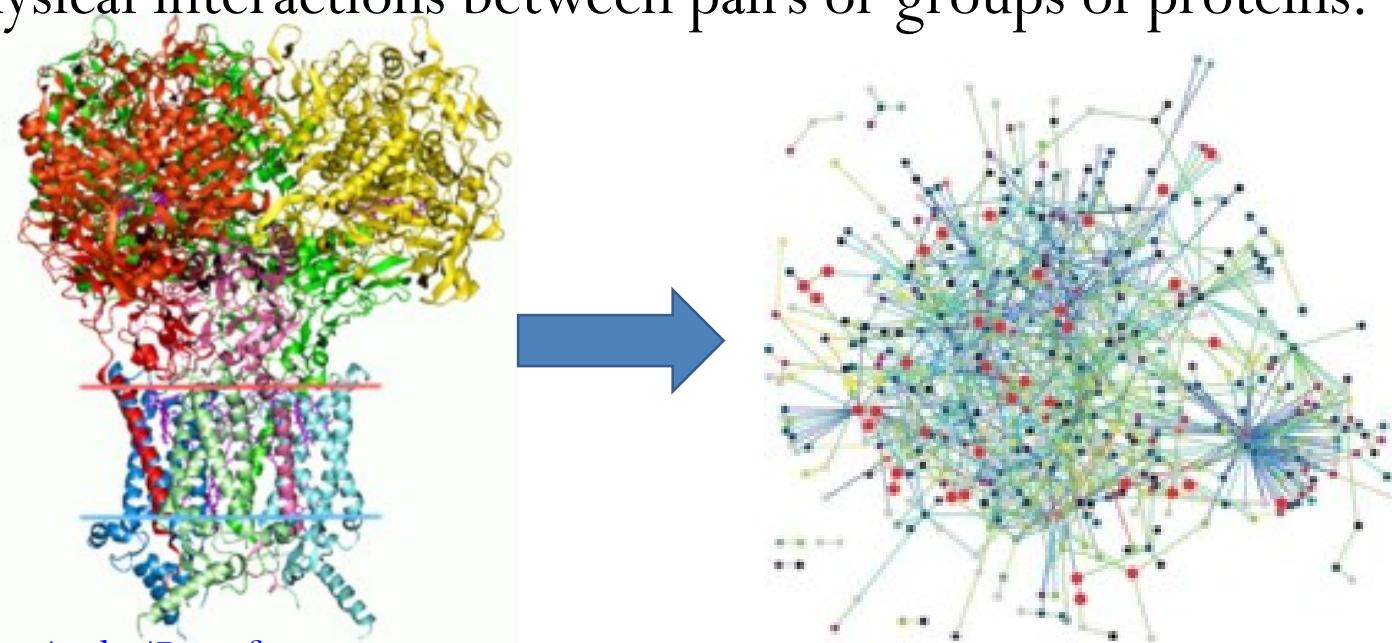
Bioinformatics data analysis

- Computers became essential in molecular biology
 - **to analyse biomedical data:** protein sequences, amino acid sequences, protein domains, protein structures etc.
- **Sequences of genetic material** are frequently used **data in bioinformatics** and are easier to manage using computers than manually.
- **DNA sequencing** is still a non-trivial problem **as the raw data** may be noisy or afflicted by weak signals.
- Algorithms have been developed for the various **data analysis** approaches to DNA sequencing.

5' ATGACGTGGGGA3'
3' TACTGCACCCCT5'

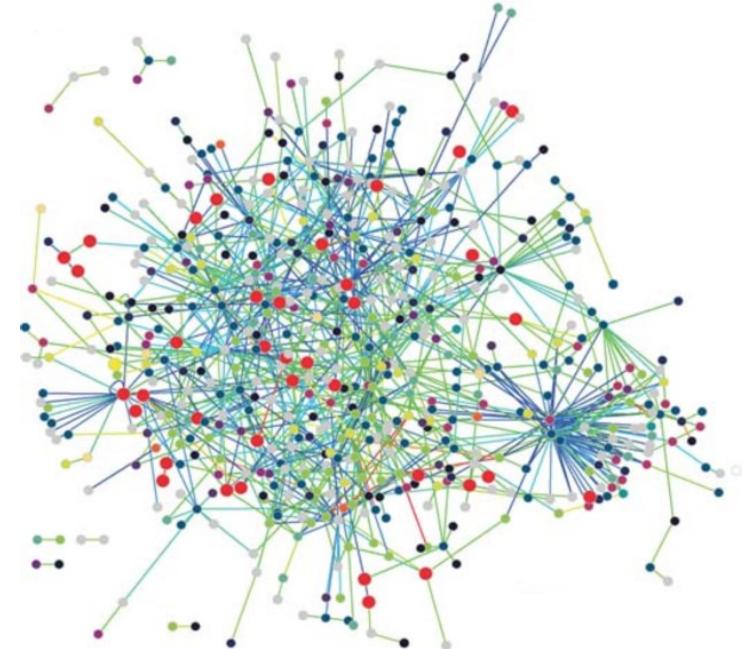
Bioinformatics data analysis

- BME: Molecular interaction networks
- Tens of thousands of three-dimensional protein structures are determined by X-ray crystallography and protein Nuclear Magnetic Resonance (NMR) spectroscopy.
- **Bioinformatics data analysis:** Protein–protein interaction identifies, predicts, and catalog physical interactions between pairs or groups of proteins.



Bioinformatics data analysis

- **Biological Network analysis**
- Study of relationships within biological networks
 - metabolic or
 - protein–protein interaction networks.
- Biological networks can be constructed from
 - a single type of molecule or entity (such as genes),
 - many different data types, such as proteins, small molecules, gene expression data.
- Abbreviation recognition – identify the long-form and abbreviation of biological terms
- Named Entity Recognition – recognizing biological terms such as gene names
- Protein–protein interaction – identify which proteins interact with which proteins from text



Bioinformatics data analysis

Other types of Bio-Interactions

- Protein-Protein Interaction
- DNA-Protein interactions
- GeneNet (Gene networks)
- Biomolecular Interaction
- Molecular interactions
- Protein and Biochemical Interactions

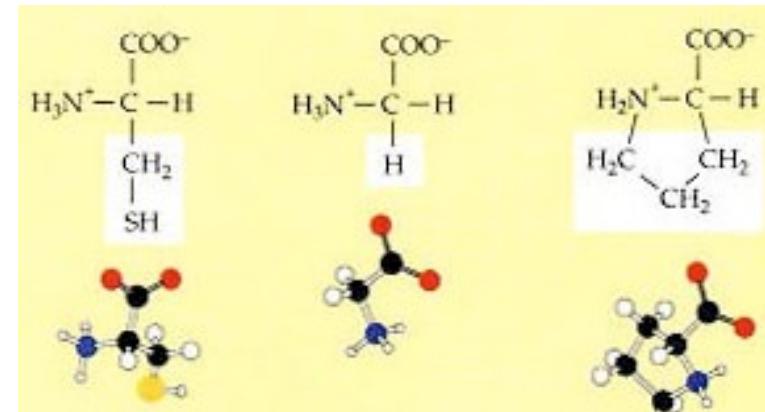
Nodes: proteins

Links: physical interactions (binding)



Bioinformatics data analysis: Visualization

- Amino Acid to Graph



- Human Hemoglobin



```
>gi|14456711|ref|NM_000558.3| Homo sapiens
hemoglobin, alpha 1 (HBA1), mRNA
ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCT
CCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGTAAGGTGGCGCGC
ACGCTGGCGAGTATGGTGGAGGCCCTGGAGAGGATGTTCTGTCCTT
CCCCACCAAGACCTACTTCCGCACTTCGACCTGAGCCACGGCTCT
GCCCAAGGTTAAGGGCACGGCAAGAAGGTGGCGACGCGCTGACCAACG
CCGTGGCGACGTGGACGACATGCCAACGCGCTGTCGCCCTGAGCGA
CCTGCACGCGACAAGCTCGGGTGGACCCGGTCAACTCAAGCTCCTA
AGCCACTGCCTGCTGGTGACCTGGCCGCCACCTCCCCGCCAGTTCA
CCCCTGCGGTGCACGCTCCCTGGACAAGTTCTGGCTTCTGTGAGCAC
CGTGTGACCTCCAAATACCCTTAAGCTGGAGCCTGGTGGCCATGCTT
CTTGCCCTTGGCCTCCCCCAGCCCCCTCCTCCCTGCACCCGT
ACCCCCGTGGTCTTGAAATAAGTCTGAGTGGCGGC
```

Bioinformatics data analysis: Databases

- KEGG (Kyoto Encyclopedia of Genes and Genomes)
 - <http://www.genome.ad.jp/kegg/>
 - Institute for Chemical Research, Kyoto University
- PathDB
 - <http://www.ncgr.org/pathdb/index.html>
 - National Center for Genomic Resources
- SPAD: Signalng PAthway Database
 - Graduate School of Genetic Resources Technology. Kyushu University.
- Cytokine Signaling Pathway DB.
 - Dept. of Biochemistry. Kumamoto Univ.
- EcoCyc and MetaCyc
 - Stanford Research Institute
- BIND (Biomolecular Interaction Network Database)
 - UBC, Univ. of Toronto

Industry 4.0

Engineering Domains to Data Science

Industrial Revolution



- Industry 1.0: transition from hand production methods to machines through the use of steam power and water power.
 - textile manufacturing, iron industry, 1760s to 1840s
- Industry 2.0 (Technological Revolution, 1870s to 1910s)
 - extensive railroad, **telegraph networks**, electrification
- Industry 3.0 (**Digital Revolution**, late 20th century)
 - **Electronics, Computer, Boolean logic, communication technologies**
- Industry 4.0 automation of traditional manufacturing and industrial practices, using modern **smart technology**.
 - **machine-to-machine communication (M2M), the internet of things (IoT), improved communication and self-monitoring, and smart machines.**
 - Mobile devices, Location detection, Human-machine interfaces, Authentication, 3D printing, Smart sensors, Big data analytics, E-Governance, Augmented reality/ wearables, Cloud computing, Data visualization

Industry 4.0

T
I
M
E
L
I
N
E

- Civil engineering
- Mechanical engineering and Industrial engineering
 - ➔ Manufacturing engineering
 - ➔ Process engineering
- Mechanical engineering and Electrical engineering
 - ➔ **Control engineering**
 - ➔ Aerospace engineering
- Mathematics and Electrical engineering
 - ➔ **Software Engineering**
 - ➔ **Computer Science and Engineering**
- Transdisciplinary (crosses many disciplines)
 - ➔ **Cybernetics**
 - ➔ **Artificial Intelligence**
 - ➔ **Data Science and Analytics**

T
I
M
E
L
I
N
E



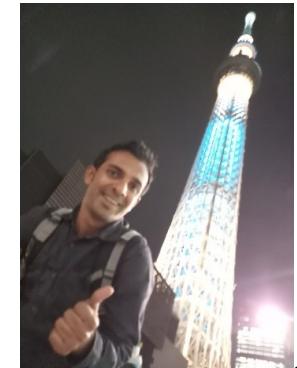
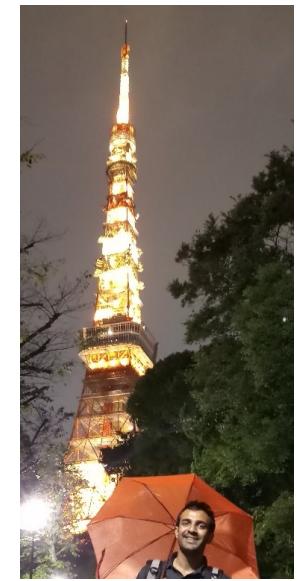
Civil engineering

- Deals with the design, construction, and maintenance of the physical and naturally built environment, including public works such as roads, bridges, canals, dams, airports, sewerage systems, pipelines, structural components of buildings, and railways
- Sub-disciplines: Coastal engineering, Construction engineering, Earthquake engineering, Environmental engineering, Forensic engineering, Geotechnical engineering, Materials science and engineering, Structural engineering, Transportation engineering, Municipal or urban engineering, Water resources engineering
- Communities: American Society of Civil Engineers, etc.



Civil engineering

- London Bridge
 - Modern bridge (1971–present), Victorian stone arch (1832–1968)
 - Medieval stone arch (1176–1832), Various wooden bridges (AD 50 – 1176)
- Tokyo Tower (1958) (332.9 meters)
 - communications and observation tower, tallest between 1958 – 1967
- Sky tree (2011 - present) (634 meters)
 - broadcasting and observation tower, tallest tower at present



1958	1967	Tokyo Tower	Japan	Tokyo	332.6 m (1,091 ft)
1967	1975	Ostankino Tower	Russia	Moscow	540.1 m (1,772 ft)
1975	2009	CN Tower	Canada	Toronto	553.33 m (1,815.4 ft)
2009	2011	Guangzhou Tower	China	Guangzhou	604 m (1,982 ft)
2011	present	Tokyo Skytree	Japan	Tokyo	634 m (2,080 ft)

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

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

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

Industrial engineering

- Concerned with the optimization of complex processes, systems, or organizations by developing, improving and implementing integrated systems of people, money, knowledge, information and equipment.
- Sub Disciplines:
 - Energy Engineering, Human Factors & Safety Engineering, Manufacturing Engineering, Operations Engineering & Management, Operations Research & Optimization, Policy Planning, Production Engineering, Quality & Reliability Engineering, Supply Chain Management & Logistics
- Völklingen Ironworks, Saarland Germany,
 - started operation in 1880s and in 1986 plant closed
 - 1994 UNESCO (World Heritage site),
 - museum focusing on making of iron,
 - preserved blast furnaces and old factory.

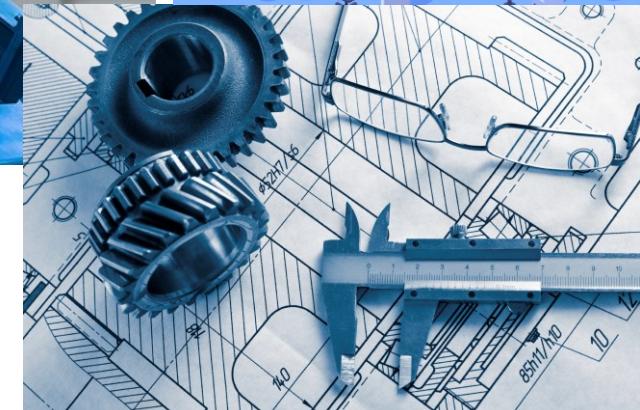
https://en.wikipedia.org/wiki/Industrial_engineering

https://en.wikipedia.org/wiki/V%C3%B6lklingen_Ironworks



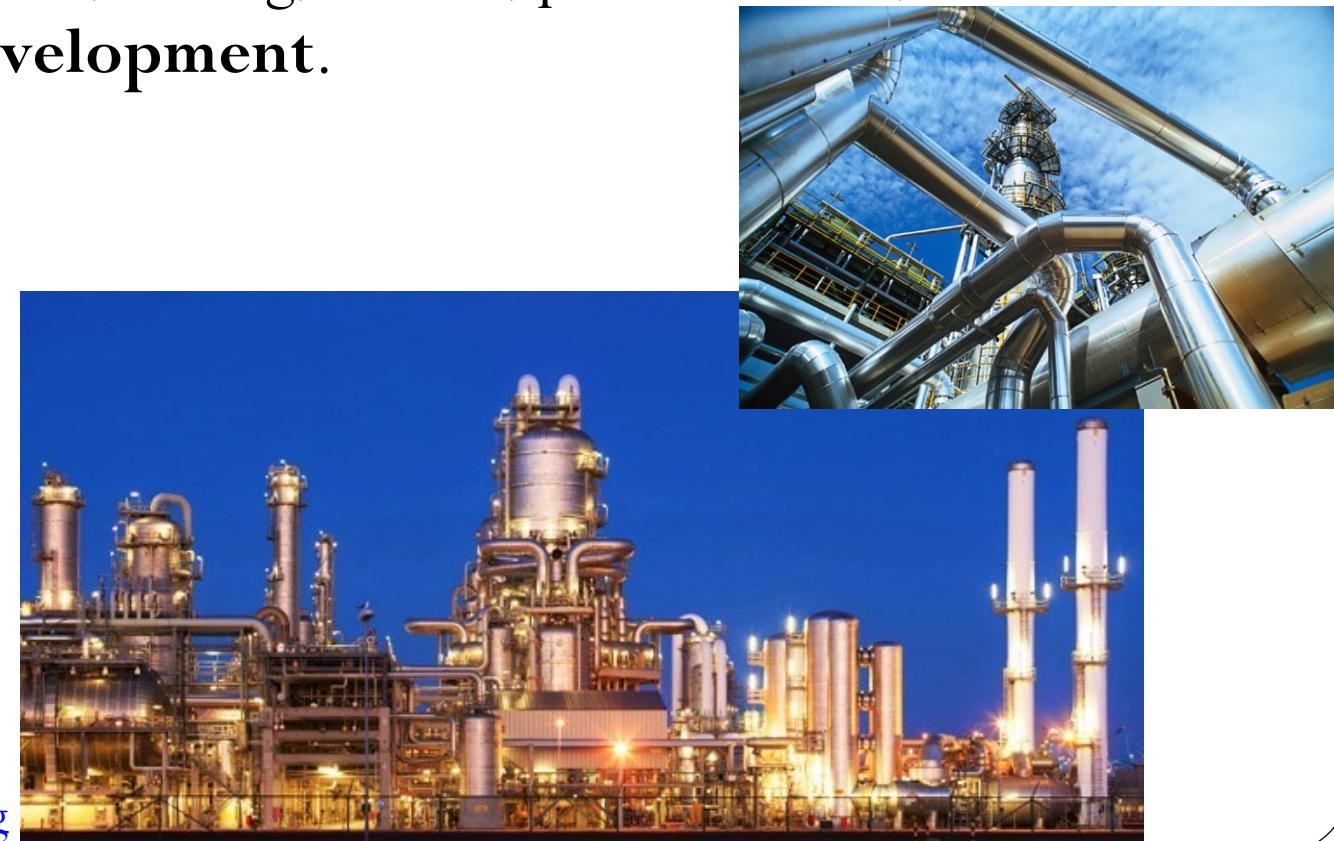
Manufacturing engineering

- To plan the practices of manufacturing; to research and to develop tools, processes, machines and equipment; and to integrate the facilities and systems for producing quality products with the optimum expenditure of capital.
- Industries where manufacturing engineers are generally employed include:
 - Aerospace industry,
 - Automotive industry,
 - Computer industry,
 - Food processing industry,
 - Textile industry,
 - Pharmaceutical industry,
 - Pulp and paper industry,
 - Toy industry



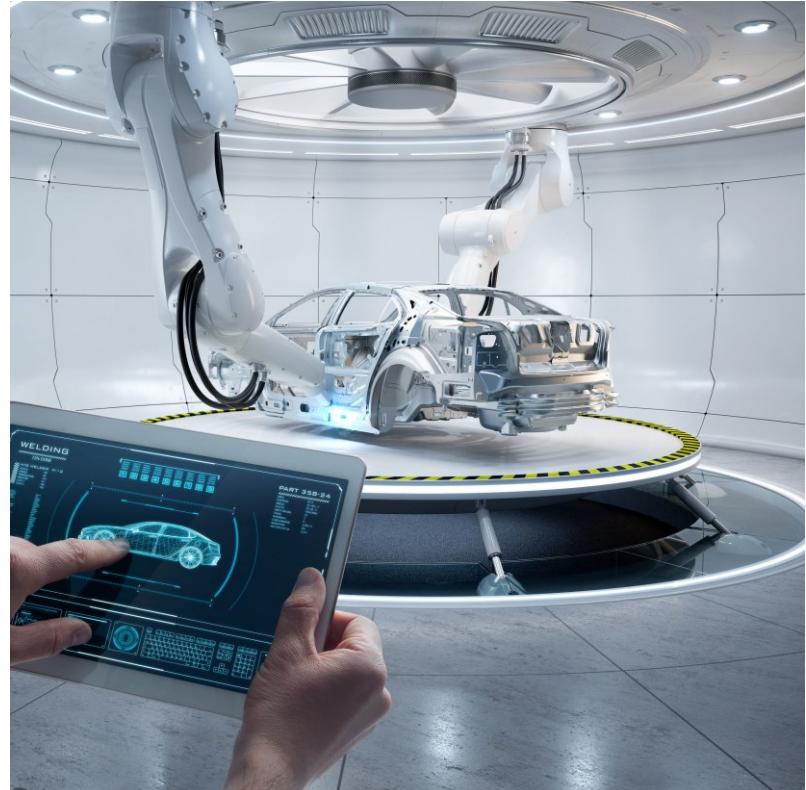
Process engineering

- Transform raw material and energy into products that are useful to society.
- Applied on a vast range of industries, such as agriculture, automotive, biotechnical, chemical, food, material development, mining, nuclear, petrochemical, pharmaceutical, and **software development**.
- Sub disciplines:
 - Process design,
 - Process control,
 - Process operations,
 - Process Economics, and
 - Process Data Analytics.



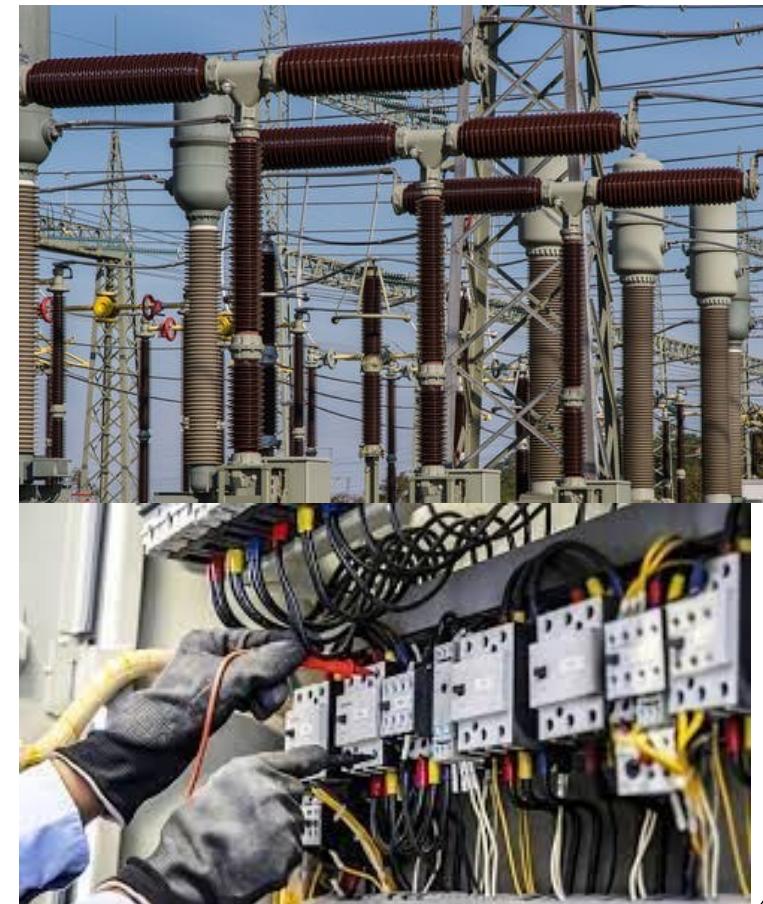
Mechanical engineering

- Combines engineering physics and mathematics principles with materials science to design, analyze, manufacture, and maintain mechanical systems
- Core areas includes:
 - mechanics, dynamics, thermodynamics,
 - materials science, **structural analysis**, and electricity.
- Tools such as
 - **Computer-aided design (CAD)**,
 - **Computer-aided manufacturing (CAM)**,
 - Industrial equipment and machinery,
 - Heating and cooling systems,
 - Transport systems, Aircraft, Watercraft, Robotics,
 - Medical devices, Weapons, and Others.
- Communities:
 - American Society of Mechanical Engineers (ASME), etc.



Electrical engineering

- Concerned with the study, design and application of equipment, devices and systems which use electricity, **electronics**, and electromagnetism
- Subdomains:
 - Power engineering,
 - Telecommunications,
 - Radio-frequency engineering,
 - Signal processing,
 - Instrumentation, and Electronics.
- Communities:
 - International Electrotechnical Commission (IEC),
 - Institute of Electrical and Electronics Engineers (IEEE) and
 - Institution of Engineering and Technology (IET)

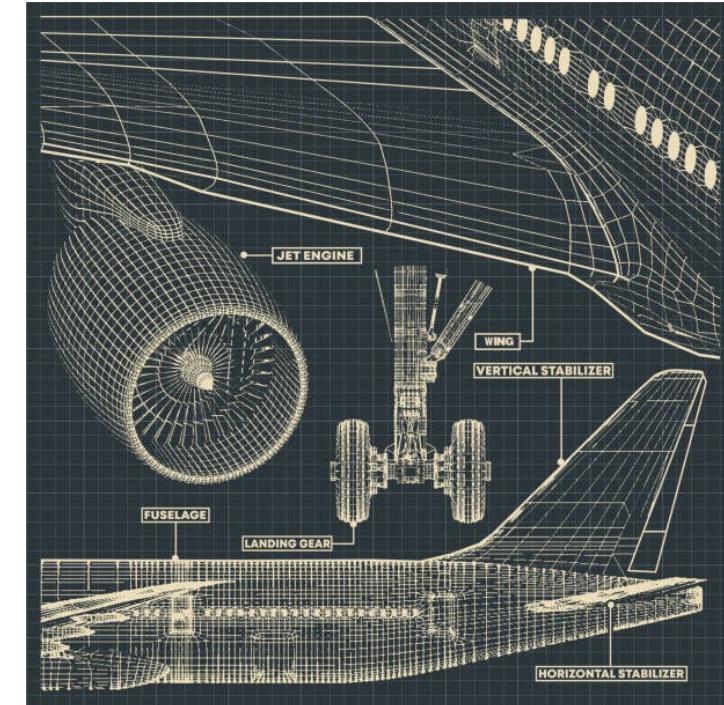


Control engineering

- Applies control theory to design equipment and systems with desired behaviors in control environments.
- Overlaps with electrical engineering and mechanical engineering
- **Sensors** and **detectors** for measuring signals.
- Without human input are **automatic control systems** (e.g. cruise control for regulating the speed of a car).
- Divisions in control theory, namely, classical and modern,
 - Classical single-input and single-output (SISO) System Design
 - Modern multiple-input and multiple-output (MIMO) System Design

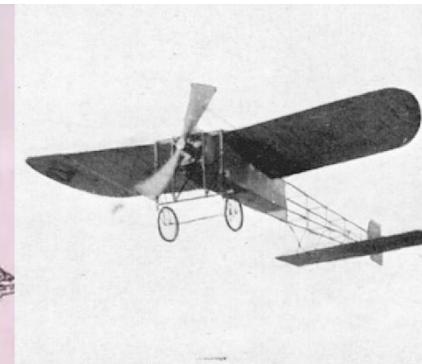
Aerospace engineering

- Concerned with the development of aircraft and spacecraft
- Overlapping branches:
 - Aeronautical engineering and Astronautical engineering
 - **Avionics** engineering deals with the **electronics** side
- Subdomains:
 - Aerodynamics,
 - Propulsion,
 - Avionics,
 - Materials science,
 - Structural analysis, and
 - Manufacturing.



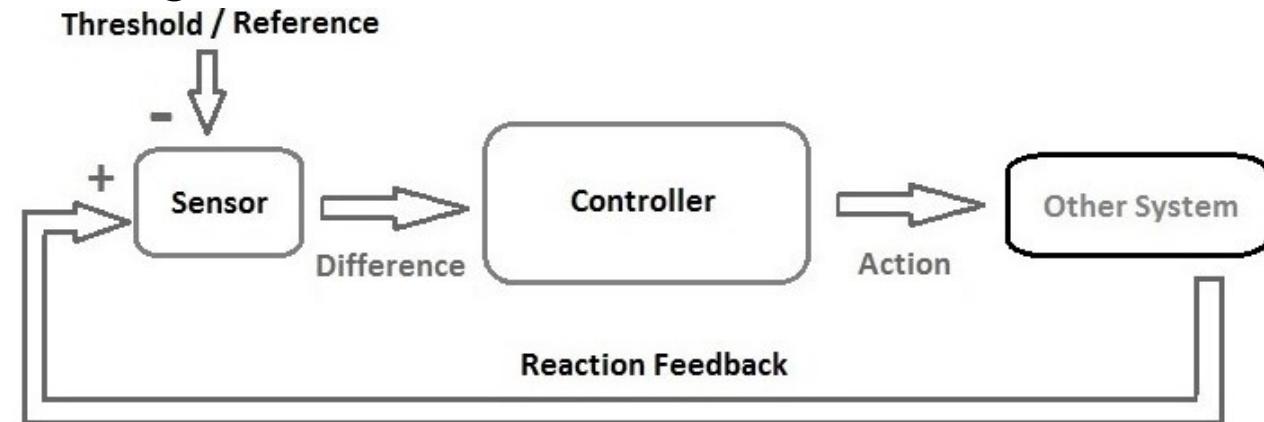
Aerospace engineering

- Shivkar Bāpuji Talpade (1864 – 1916)
 - made first unmanned aircraft in 1895
 - named *Marutsakhā*, derived from the Sanskrit *Marut* ('air' or 'stream') and *sakhā* ('friend') which together mean 'Friend of wind'
 - inspired from Vimāna, ancient flying-machines
 - D. K. Kanjilal's 1985 “*Vimana in Ancient India: Aeroplanes Or Flying Machines in Ancient India*”,
- In December 1903, the Wright Brothers
 - first sustained,
 - controlled flight of a powered,
 - heavier-than-air aircraft,
 - lasting 12 seconds.



Cybernetics (Trans-disciplinary)

- Cybernetics has its origins around 1940s from anthropology, mathematics, neuroscience, psychology, and engineering. Initial meetings such as Macy Conferences and Ratio Club.
- It is prominent during the 1950s and 1960s. It is a precursor to fields such as Computing, Artificial Intelligence, Cognitive Science, Complexity science, and Robotics.
- Concepts to explore system's - structures, constraints, and possibilities, learning, cognition, adaptation, social control, emergence, and connectivity
- Intersection of the fields of control systems, electrical network theory, mechanical engineering, neuroscience, art, architecture and design, biology, computer science, engineering, earth system science, law, management, and mathematics.
- feedback - the outcomes of actions are taken as inputs for further action.

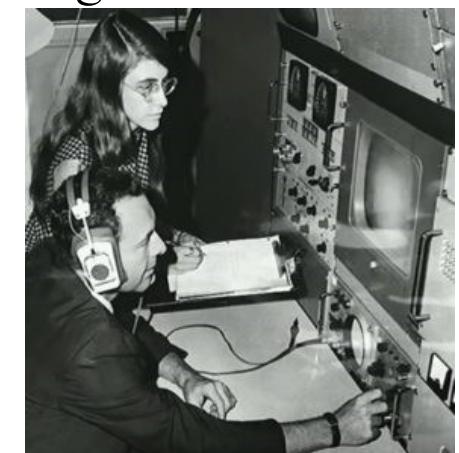


Computer Science Engineering (CSE)

- Comprises scientific and engineering aspects of computing.
- In Europe, it is also an often term used to translate as Engineering Informatics
- **Systems subjects:** Programming, Algorithms, Data structures, Computer architecture, Operating systems, Computer networks, Parallel computing, Digital logic, Computer graphics, Software engineering, Database systems etc.
- **Theory subjects:** Theoretical computer science, Numerical methods, Machine learning, Programming theory etc.
- **Mathematical courses:** Discrete mathematics, Mathematical analysis, Linear algebra, Probability, Statistics, Optimization techniques etc.
- **Modern emerging computing:** Cloud computing, Computer vision, Computer Security, Data science, Robotics, Bio-inspired computing, Computational biology, Autonomic computing and Artificial intelligence.

Software engineering and AI

- Digital computers appeared in the early 1940s
 - "Stored program architecture"
 - Von Neumann architecture
 - division between "hardware" and "software"
- Programming languages started to appear in the early 1950s
 - Fortran, ALGOL, PL/I, and COBOL
- Systematic application of engineering approaches to the development of software.
- "Software Engineering" 1965 issue of COMPUTERS and AUTOMATION magazine
- Margaret Hamilton coined the term
 - "software engineering"
 - during the Apollo missions
- Communities: ICSE, FSE, ICSME, ASE, SANER



Artificial Intelligence

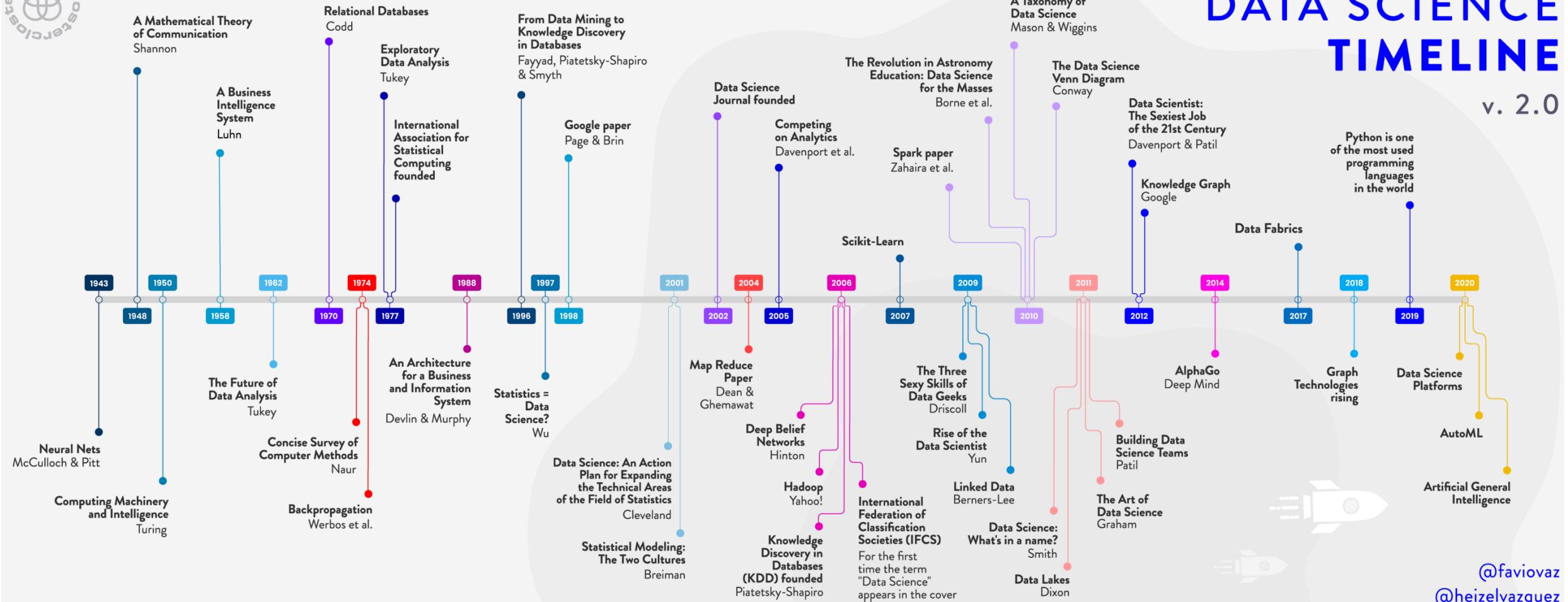
- Founded in 1956 (Dartmouth Research Project on Artificial Intelligence).
- Ancient History shows that AI was an active field of studies since ancient times.
- Simulating the brain, modeling human problem solving, formal logic, large databases of knowledge and imitating human behavior
- Search, Optimization, Formal logic,
- Artificial Neural Networks, Deep learning,
- Classifiers, Reasoning, Knowledge Representation,
- Planning, Learning, Robotics,
- Natural Language Processing, Computational Social Scienc, etc.

Artificial Intelligence

- (1837) Analytical Engine, (1925) Houdina Radio Control,
- (1946-1960) ENIAC heralds the dawn of Computing,
- (1946-1960) Turing Test, (1959) Machine Learning,
- (1964) Eliza Chatbot, (1966) Shakey general purpose mobile robot,
- (1960-74) Expert Systems,
- AI Winter (1974-80), AI Spring (1980-87), AI Winter (1987-93),
- (1993) Apriori algorithm: Association Rule Mining,
- (1997) Deep Blue ends Human Supremacy in Chess,
- (1998) Deep Space, (2005) Driverless Cars,
- (2011) IBM Watson beaten Human in Jeopardy,
- (2016) Alpha Go

Data Science

DATA SCIENCE TIMELINE v. 2.0



The Roots of Data Science by Favio Vázquez

<https://towardsdatascience.com/the-roots-of-data-science-77c71115229>

@faviovaz

@heizelvazquez



Data Science

A Mathematical Theory
of Communication
Shannon

A Business
Intelligence
System
Luhn

Relational Databases

Codd

Exploratory
Data Analysis
Tukey

International
Association for
Statistical
Computing
founded

From Data Mining to
Knowledge Discovery
in Databases

Fayyad, Piatetsky-Shapiro
& Smyth

8

Data Science
Journal founded

Com
on AI
Dave

1943

1950

1948

1958

1962

1970

1974

1988

1996

1997

1998

2001

2002

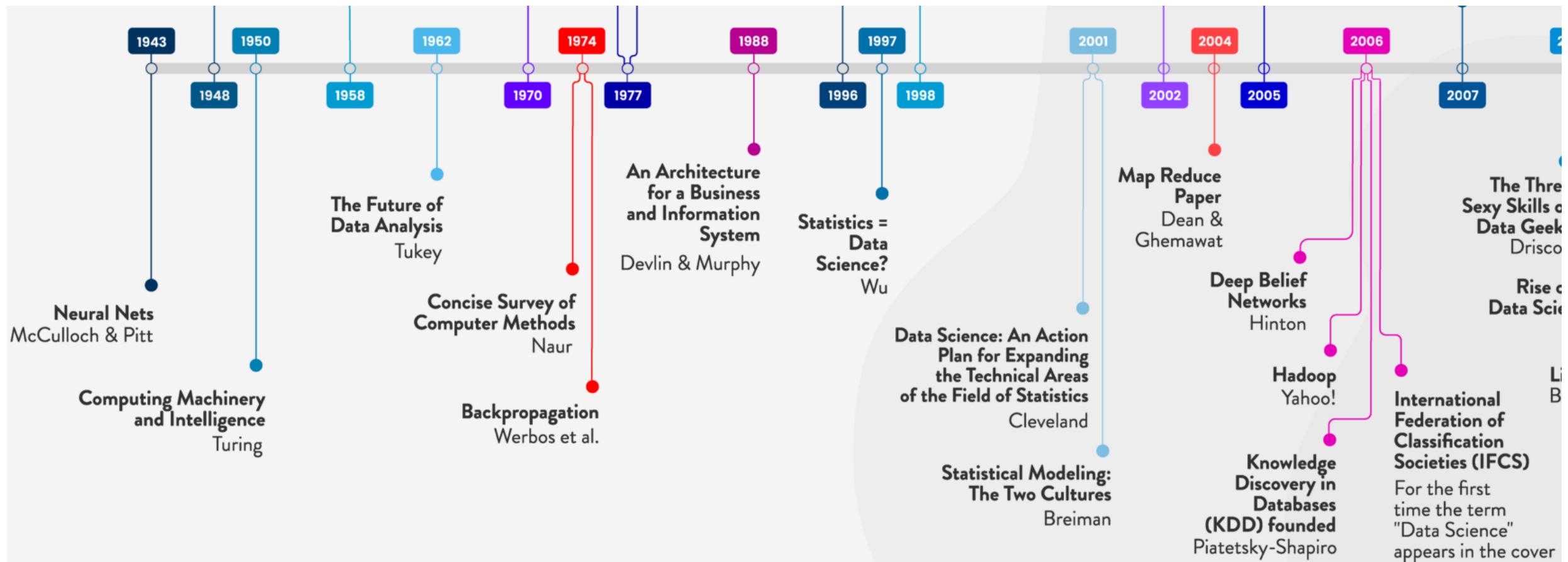
2004

2005

The Roots of Data Science by [Favio Vázquez](#)

<https://towardsdatascience.com/the-roots-of-data-science-77c71115229>

Data Science



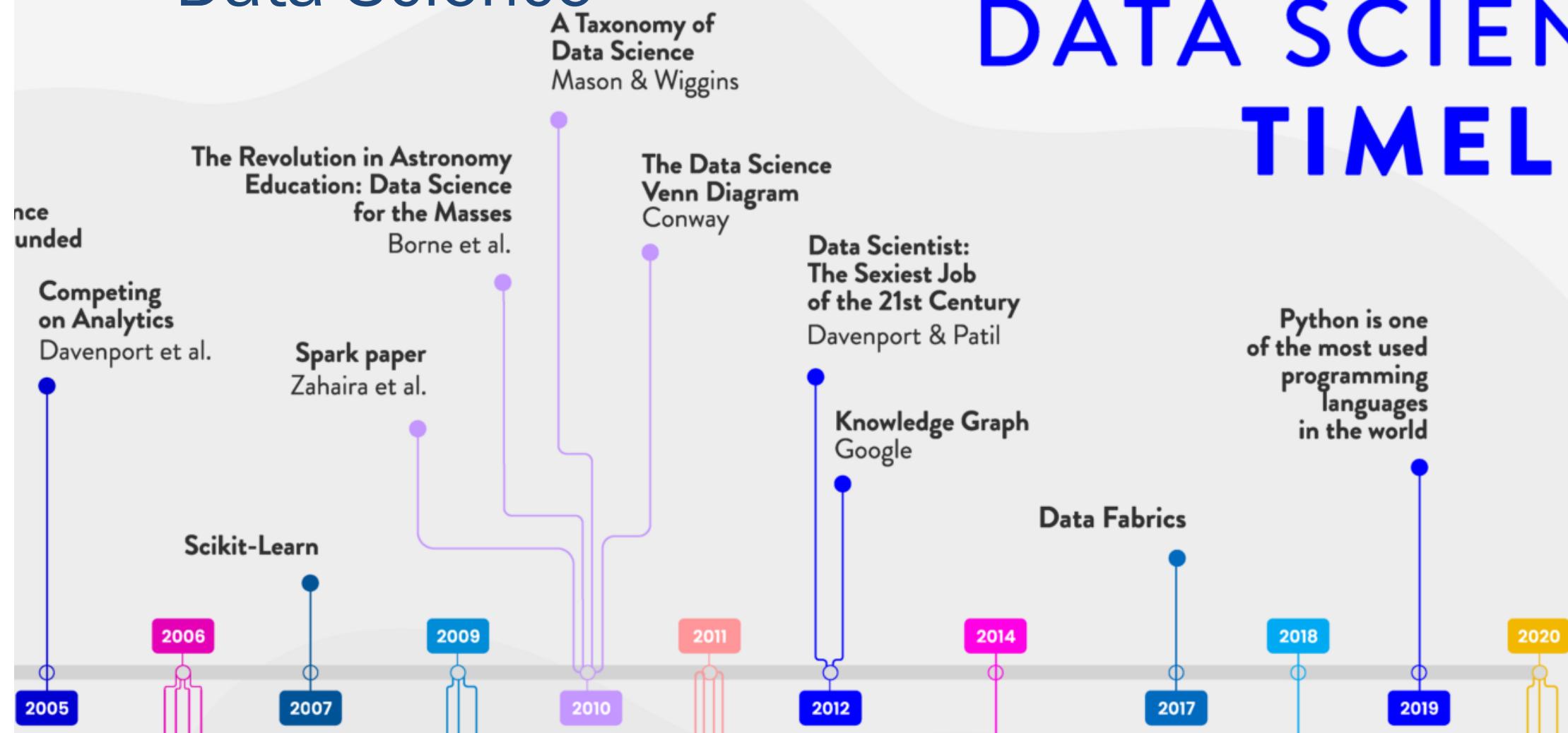
The Roots of Data Science by [Favio Vázquez](#)

<https://towardsdatascience.com/the-roots-of-data-science-77c71115229>

Data Science

DATA SCIENCE TIMELINE

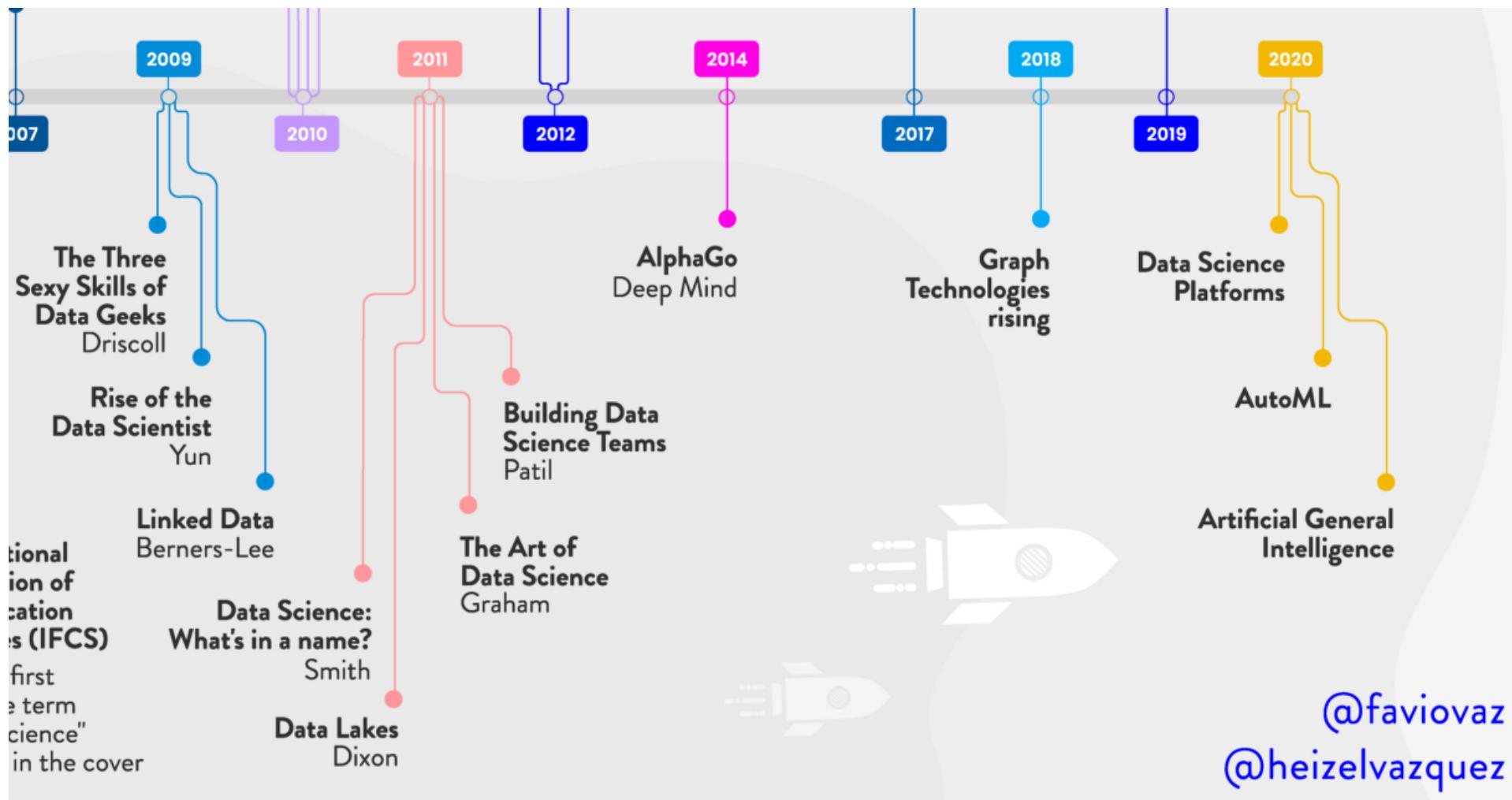
v. 2.0



The Roots of Data Science by [Favio Vázquez](#)

<https://towardsdatascience.com/the-roots-of-data-science-77c71115229>

Data Science



The Roots of Data Science by [Favio Vázquez](#)

<https://towardsdatascience.com/the-roots-of-data-science-77c71115229>

תודה רבה

Hebrew

Danke

German

Merci

French

Grazie

Italian

Gracias

Spanish

Obrigado

Portuguese

Ευχαριστώ

Greek

Спасибо

Russian

ধন্যবাদ

Bangla

ಧನ್ಯವಾದಗಳು

Kannada

ధన్యవాదాలు

Telugu

ਧੰਨਵਾਦ

Punjabi

धन्यवादः

Sanskrit

Thank You

English

நன்றி

Tamil

മന്ത്രി

Malayalam

આમાર

Gujarati

شُكْرًا

Arabic

多謝

Traditional Chinese

多谢

Simplified Chinese

ありがとうございました

Japanese

ຂອບຄຸມ

Thai

감사합니다

Korean

<https://sites.google.com/site/animeshchaturvedi07>