Artificial intelligence and health informatics

Dr Jean-Charles DUFOUR

Last update : 26 october 2021

SESSTIM, Faculté des Sciences Médicales et Paramédicales, Aix-Marseille Université, Marseille, France
https://sesstim.univ-amu.fr/

Outline

- AI (in the field of health informatics)
  - Definition
  - Brief history
  - AI typologies
  - Potential for AI in healthcare and public health
  - Challenges
- Data for AI
  - What are data ?
  - Notions of information systems interoperability
  - Data reusability

Definition

"The capacity of computers or other machines to exhibit or ***simulate intelligent behavior****; the field of study concerned with this"* (Oxford English Dictionary)

"Systems that ***mimic cognitive functions*** *generally associated with human attributes such as learning, speech and problem solving"* (Russel & Norvig's book Artificial Intelligence)

Other definitions are focused on AI goal types, tasks, applications and methods (In their recent review Collins et al. list 28 definitions !)

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Definition

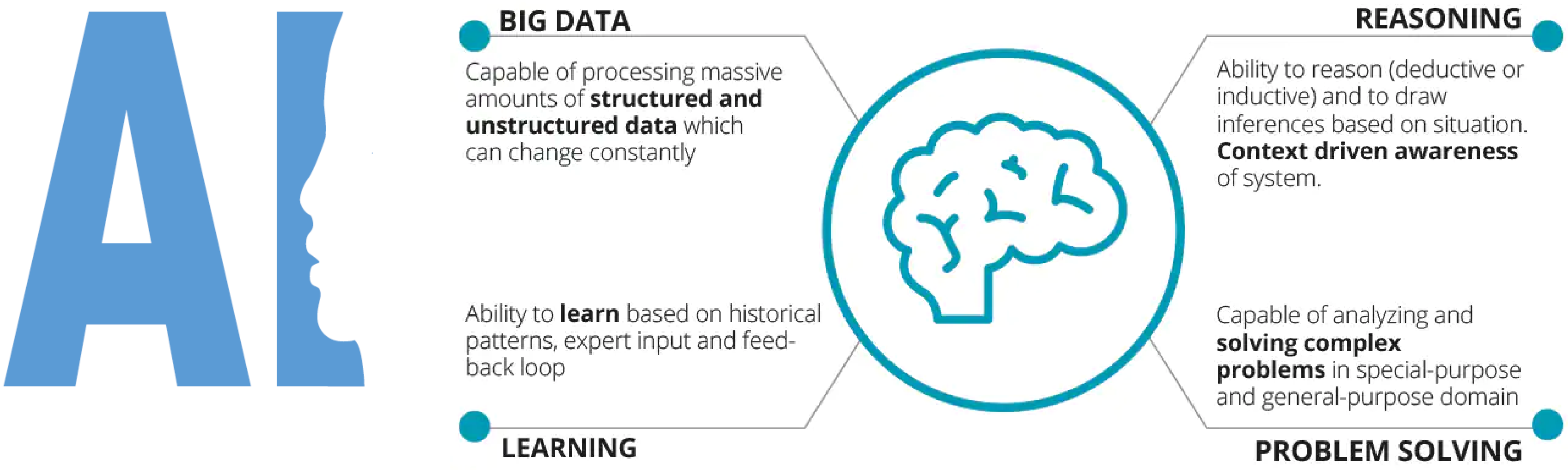
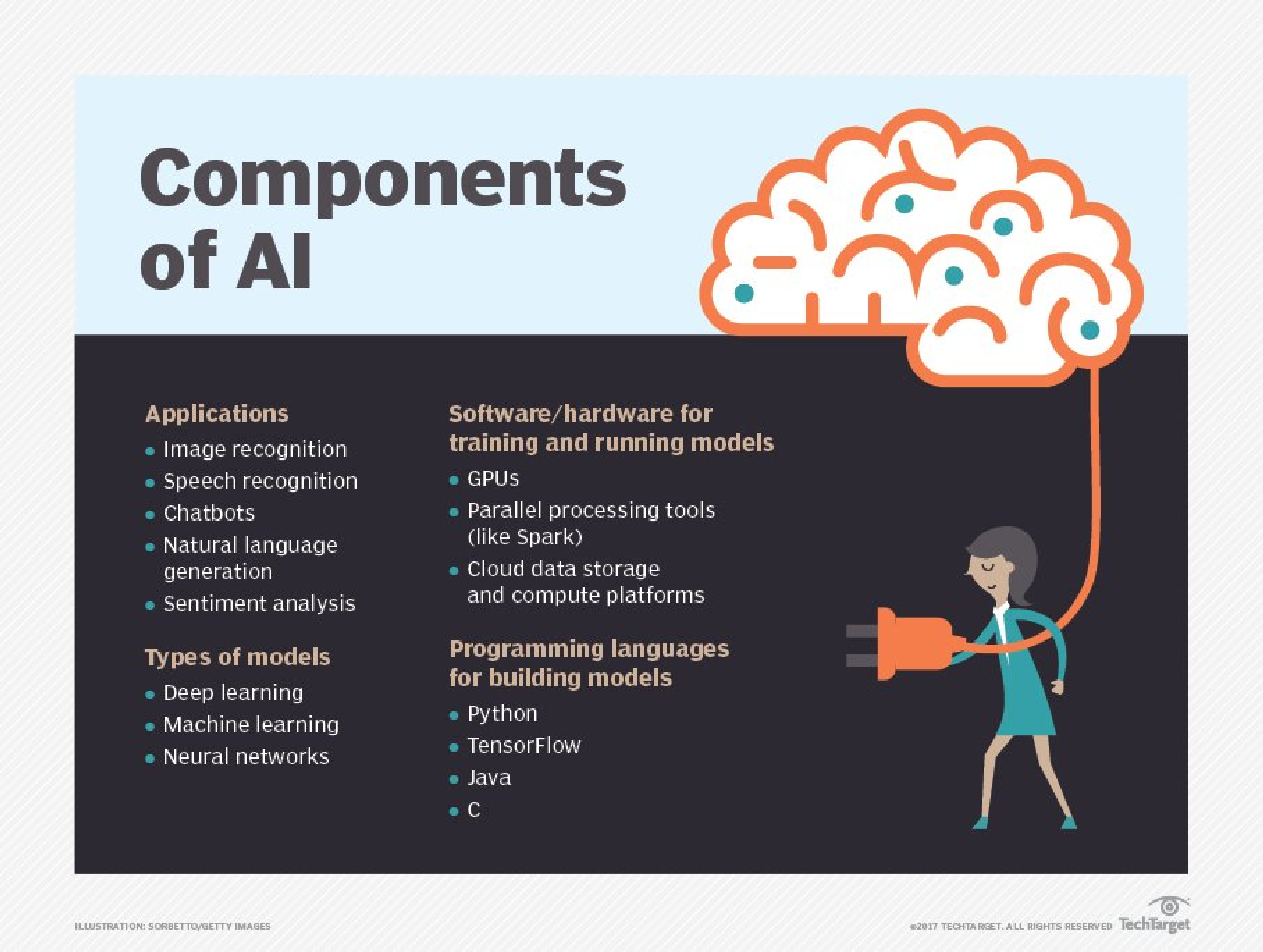


Image from deloitte.com

SESTIM



Components of AI
Applications
Image recognition
Speech recognition
Chatbots
Natural language generation
Sentiment analysis
Types of models
Deep learning
Machine learning
Neural networks
Software/hardware for training and running models
GPUs
Parallel processing tools (like Spark)
Cloud data storage and compute platforms
Programming languages for building models
Python
TensorFlow
Java
C
ILLUSTRATION: SORBETTO/GETTY IMAGES
©2017 TECHTARGET. ALL RIGHTS RESERVED
TechTarget

Definition

Image from online.king.edu

*"A **transdisciplinary** study of the **data** flow and processing into more abstract forms such as **information**, **knowledge**, and **wisdom** along with the associated systems needed to synthesize or develop decision support systems for the purpose of helping the **healthcare management processes** achieve better outcomes in healthcare delivery."* (Wan T. & Gurupur V.)

[2020, Understanding the difference between healthcare informatics and healthcare data analytics in the present state of health care management, https://doi.org/10.1177/2333392820952668]

Health Informatics ⟷

(Health Information System)

Definition

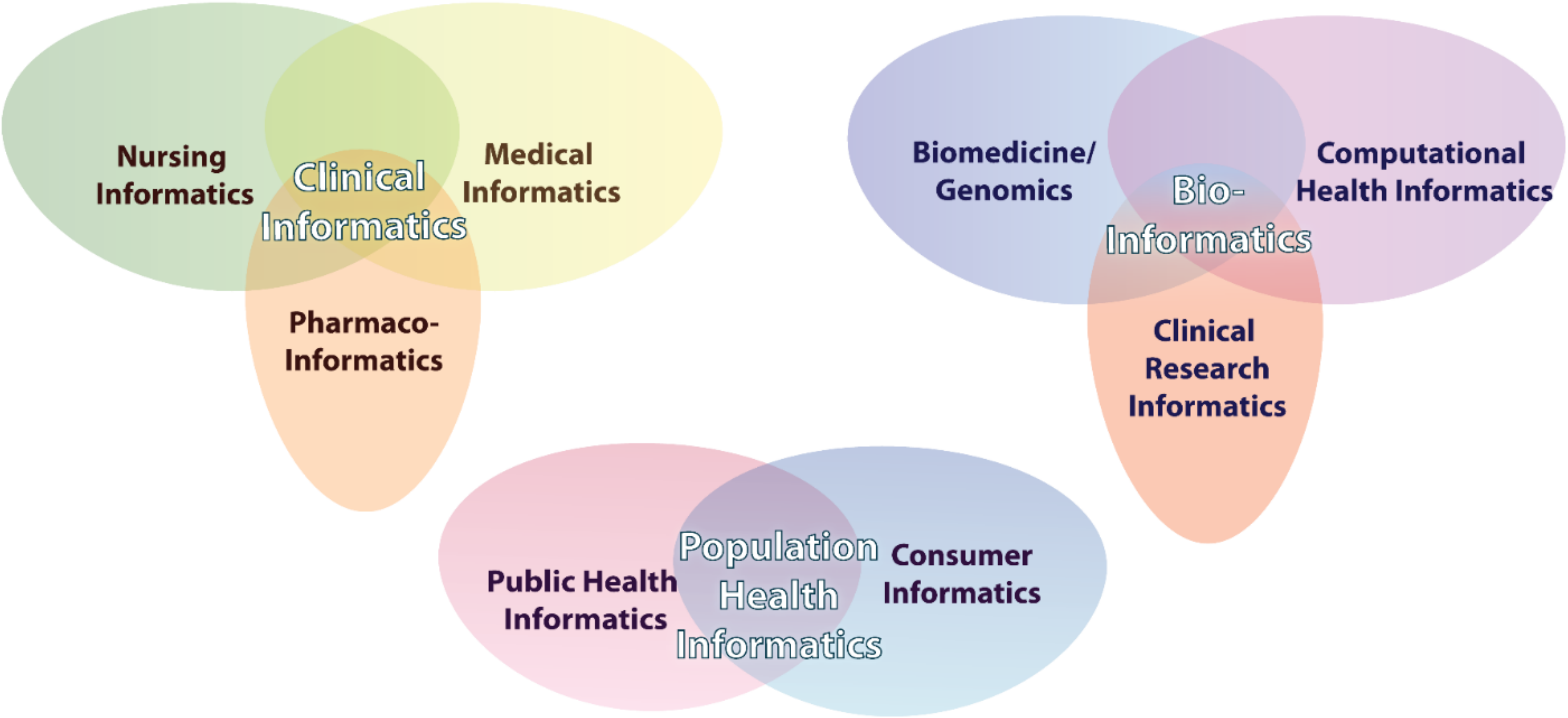
Image from online.king.edu

*"A **transdisciplinary** study of the **data** flow and processing into more abstract forms such as **information**, **knowledge**, and **wisdom** along with the associated systems needed to synthesize or develop decision support systems for the purpose of helping the **healthcare management processes** achieve better outcomes in healthcare delivery."* (Wan T. & Gurupur V.)

[2020, Understanding the difference between healthcare informatics and healthcare data analytics in the present state of health care management, https://doi.org/10.1177/2333392820952668]

Health Informatics	≠	**Health IT**
(Health Information System)		(Health InformationTechnologies)
How and why behing health IT		Use of technology in health care

Health Informatics

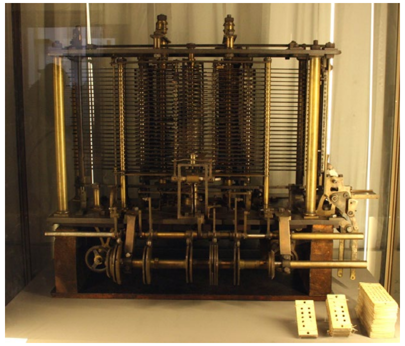


This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552 EUUSEHEALTHWORK

Brief (pre)history

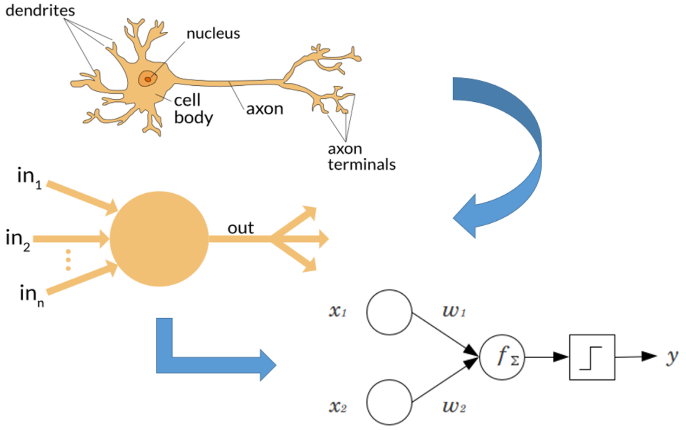
1642 : First mechanical calculating machine built by B. Pascal

see https://youtu.be/GX4RQK__fQc and https://youtu.be/hSl2WFfCTD8 (in French language) if you're curious

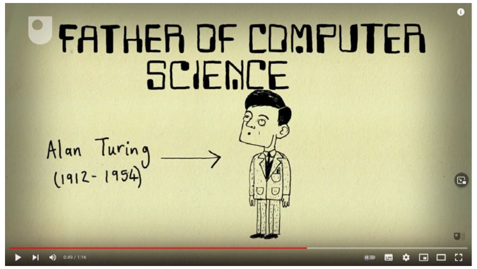


1837 : First design for a programmable machine by C. Baddage

Brief history



1943 : Foundations of neural networks by W. McCulloch and W. Pitts



1950 : the Turing test and Turing machine by A. Turing

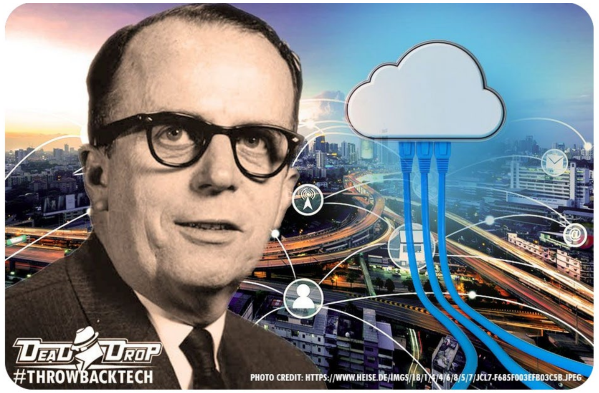
see https://youtu.be/TryOC83PH1g if you're curious about Chinese Room *vs* Turing test

Image from https://pemey.medium.com/qui-est-ce-jeune-riant-avec-des-chaussettes-hautes-en-plein-%C3%A9t%C3%A9-5060f99ac459

1955 : "Dartmouth Summer Research Project on Artificial Intelligence" by J. McCarthy, M. Minsky, N. Rochester and C. Shannon

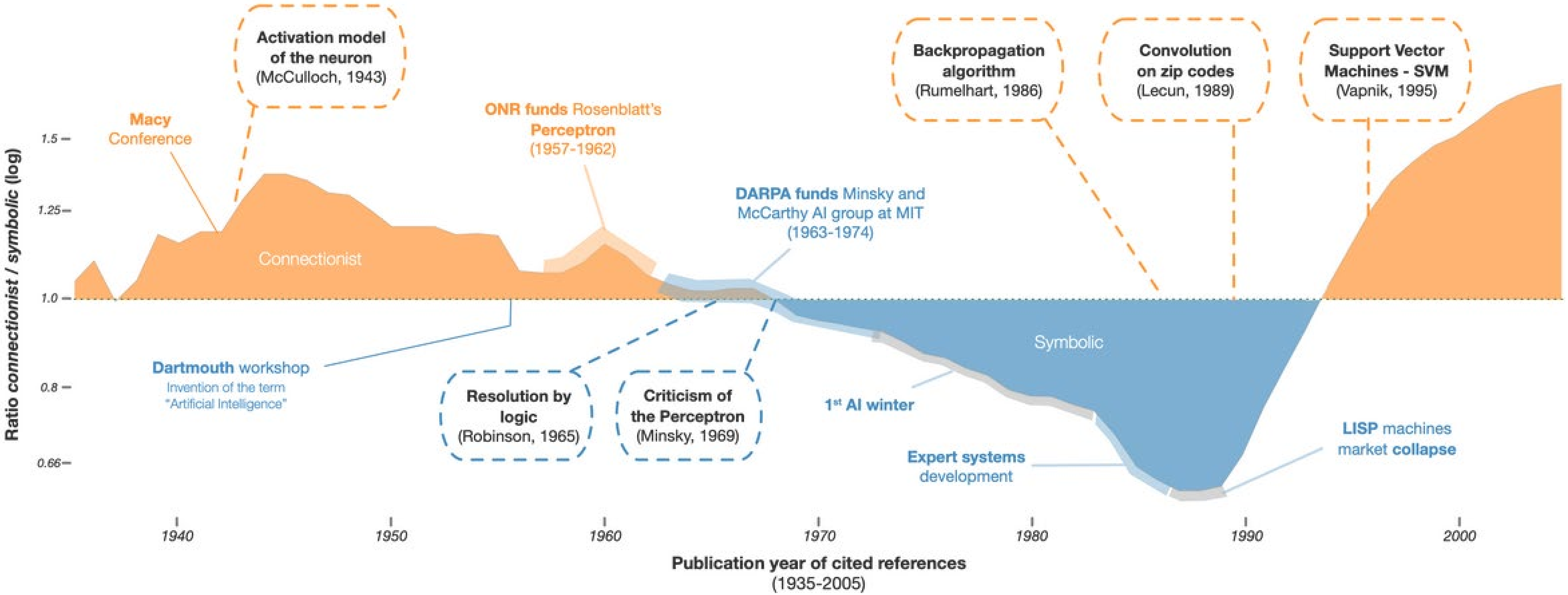
see https://ojs.aaai.org//index.php/aimagazine/article/view/1904 for the program !!!

Brief history



"The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today." (1960 J.C.R. Licklider)

Brief history : Connectionist and symbolic approaches



From D. Cardon, JP Cointet, A. Mazières. Neurons spike back - The invention of inductive machines and the controverse of Artificial Intelligence.
https://neurovenge.antonomase.fr/

Brief history : health informatics expert systems

1971 : HELP – (*Salt Lake City*)
by Wamer HR, Olnstead CM, Rutherford BD

https://core.ac.uk/download/pdf/276276919.pdf

1972-82 : INTERNIST-1 and its successor, Quick Medical Reference (QMR) – (*Pittsburgh*)
by HE Pople, JD Myers, RA Miller

http://www.skateboardingalice.com/papers/1986_Miller.pdf

1975 : MYCIN – (*Stanford*)
by EH Shortliffe and BG Buchanan

https://doi.org/10.1016/0025-5564(75)90047-4

1984 : DXplain – (*Massachusetts General Hospital*)
by Barnett GO, Cimino JJ et al.

http://www.mghlcs.org/projects/dxplain/

…and others… (see https://www.clinfowiki.org/wiki/index.php/Timeline_of_the_Development_of_Clinical_Decision_Support)

Brief history

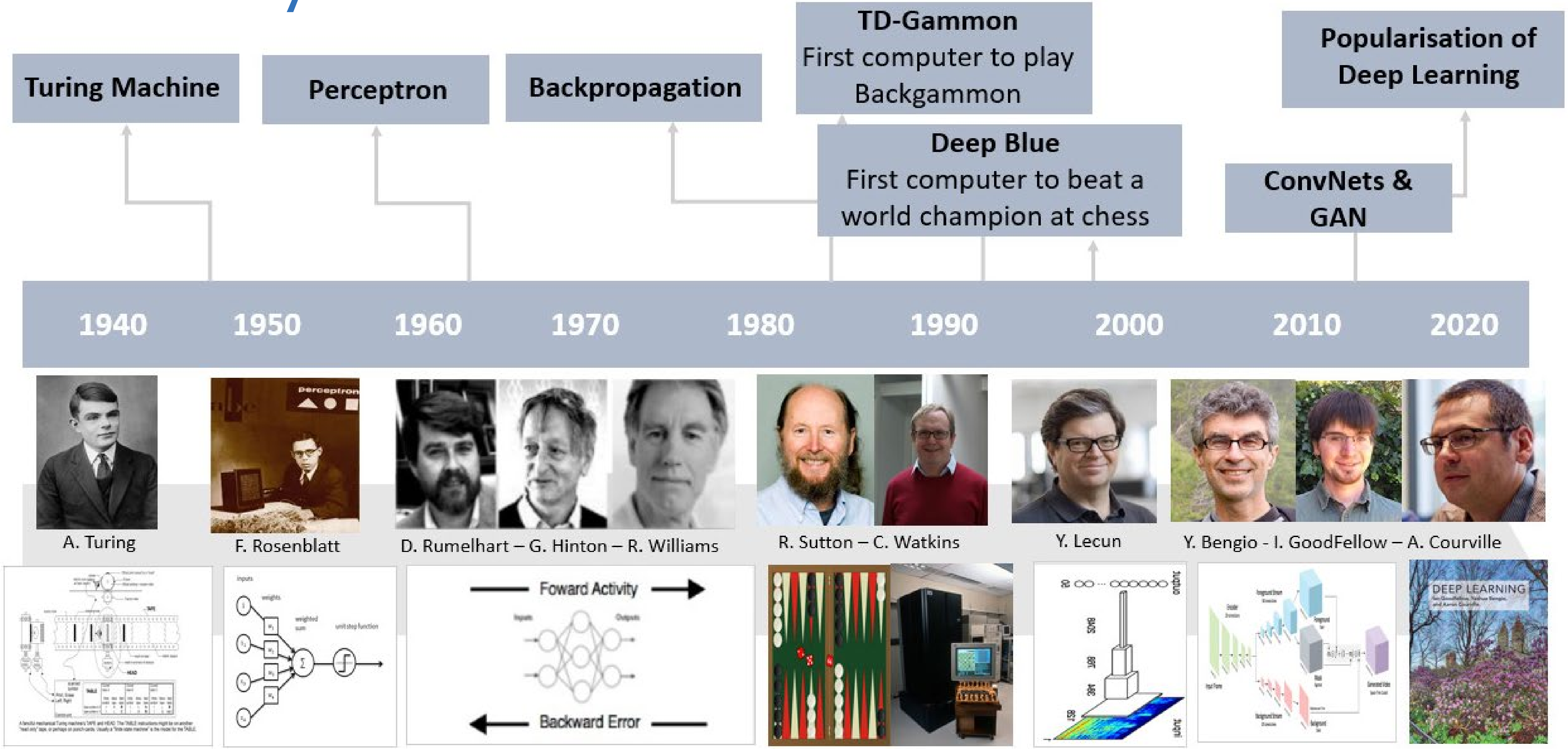


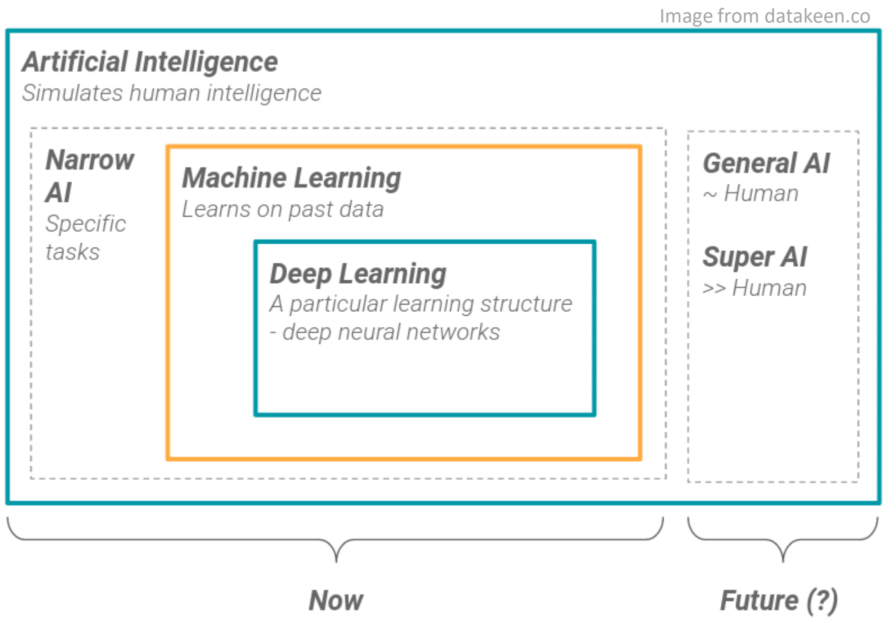
Image from Mohamed Hanini. The State of Artificial Intelligence and Its Applications. https://koiosintelligence.ca/the-state-of-artificial-intelligence-and-its-applications/

IA typologies

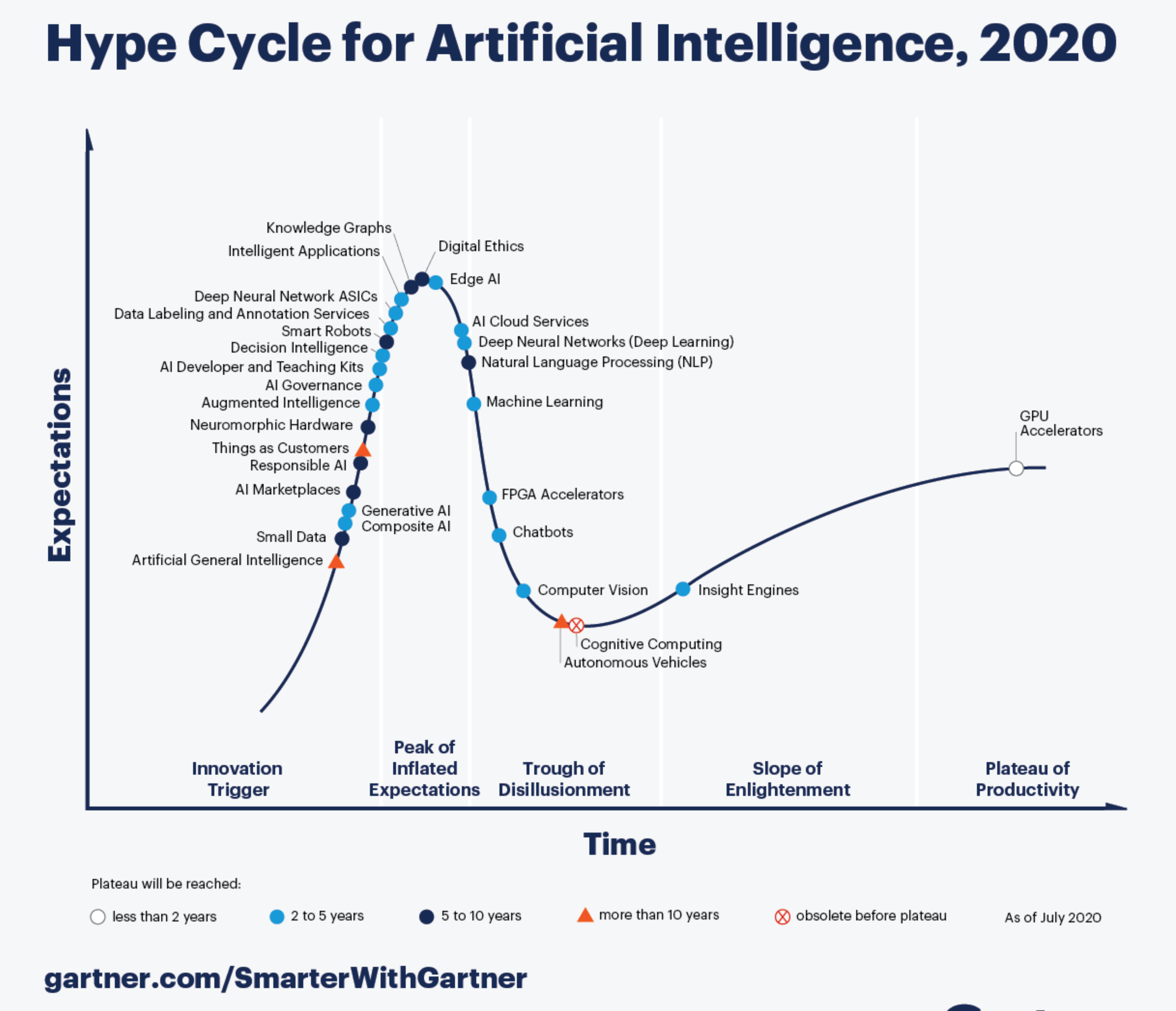
Several possible typologies:

- « Coverage ambitions »
- Objectives pursued
- Applications focused
- Methods used
- …

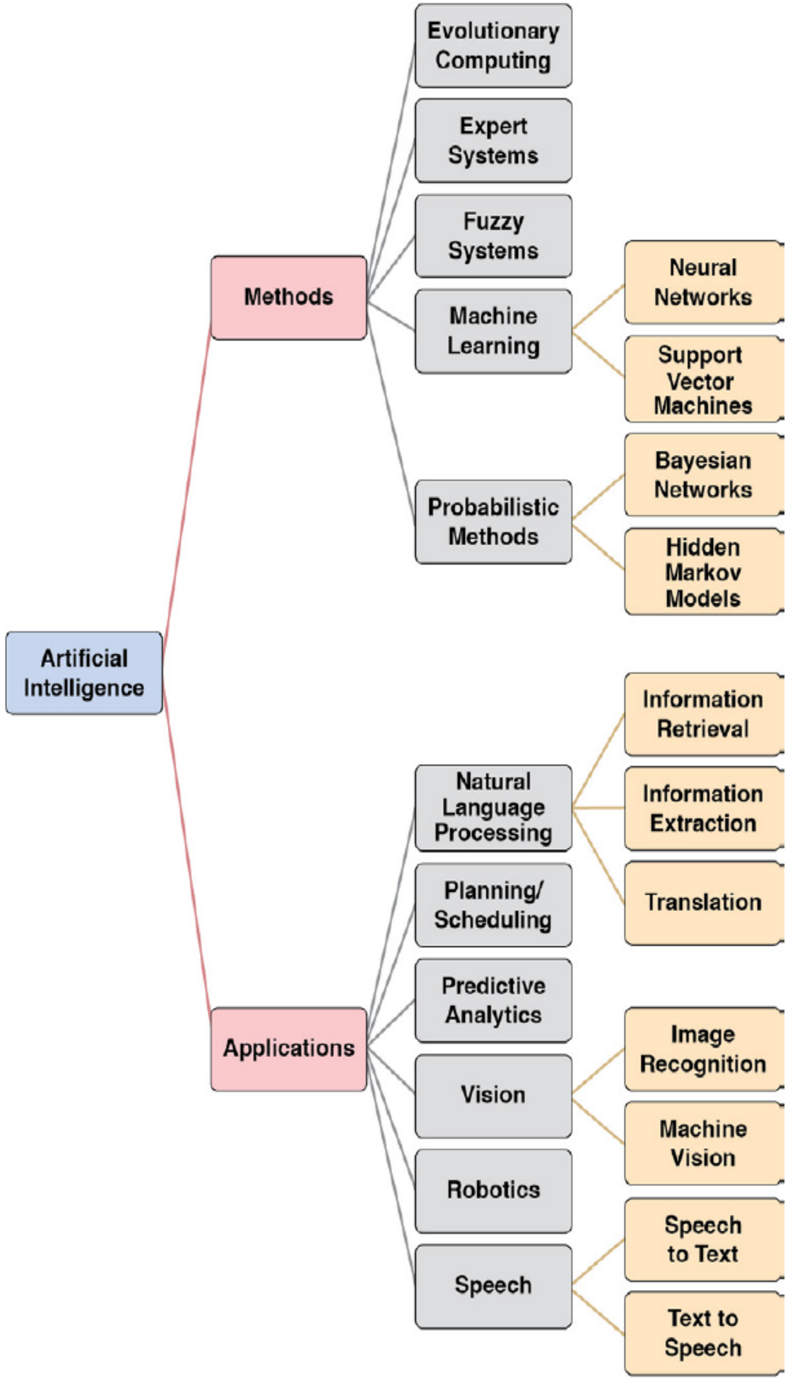
Super, General and Narrow AI



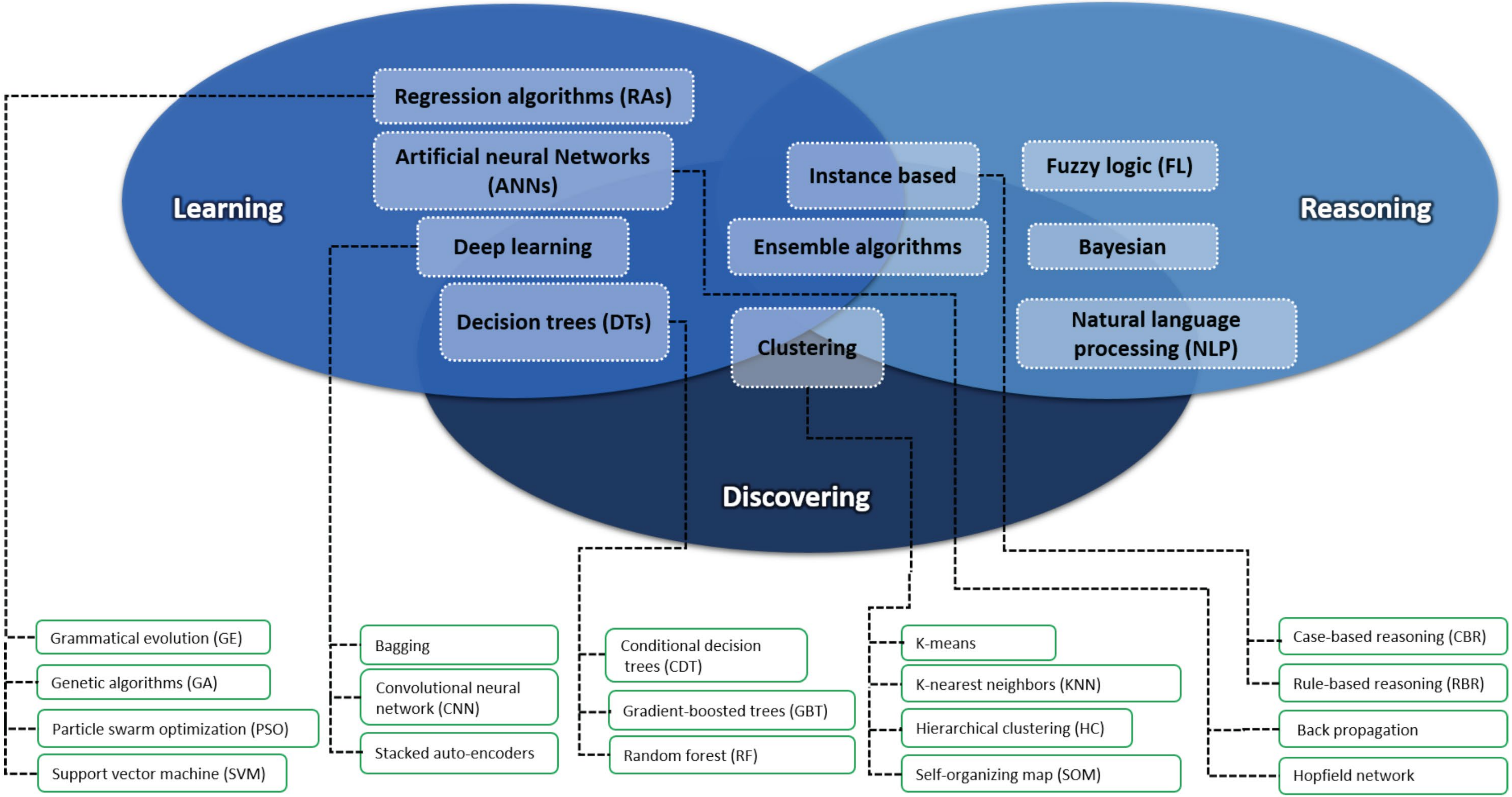
- Super AI :
  - Machines that are much more smarter than humans
  - Fictional, singularity theory
- General or strong AI :
  - Machines that would be able to apply apply knowledge and skills in different contexts
  - A small research community exist (Deepmind, Cyc, OpenAI,…)
- Narrow AI
  - Algorithm specialized at a single task
  - Many systems already exists (playing chess, driving car, face recognition, surgical robots, Skin or X-ray images analysis, …


Hype Cycle for Artificial Intelligence, 2020
Expectations
Knowledge Graphs
Intelligent Applications
Digital Ethics
Edge AI
Deep Neural Network ASICs
Data Labeling and Annotation Services
Smart Robots
Decision Intelligence
AI Developer and Teaching Kits
AI Governance
Augmented Intelligence
Neuromorphic Hardware
Things as Customers
Responsible AI
AI Marketplaces
Generative AI
Composite AI
Small Data
Artificial General Intelligence
AI Cloud Services
Deep Neural Networks (Deep Learning)
Natural Language Processing (NLP)
Machine Learning
FPGA Accelerators
Chatbots
Computer Vision
Insight Engines
Cognitive Computing
Autonomous Vehicles
GPU Accelerators
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Time
Plateau will be reached:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau
As of July 2020
gartner.com/SmarterWithGartner
Source: Gartner

From : "*Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*". National Academy of Medicine. https://nam.edu/artificial-intelligence-special-publication/



From Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018;20(5):e10775. doi: 10.2196/10775.



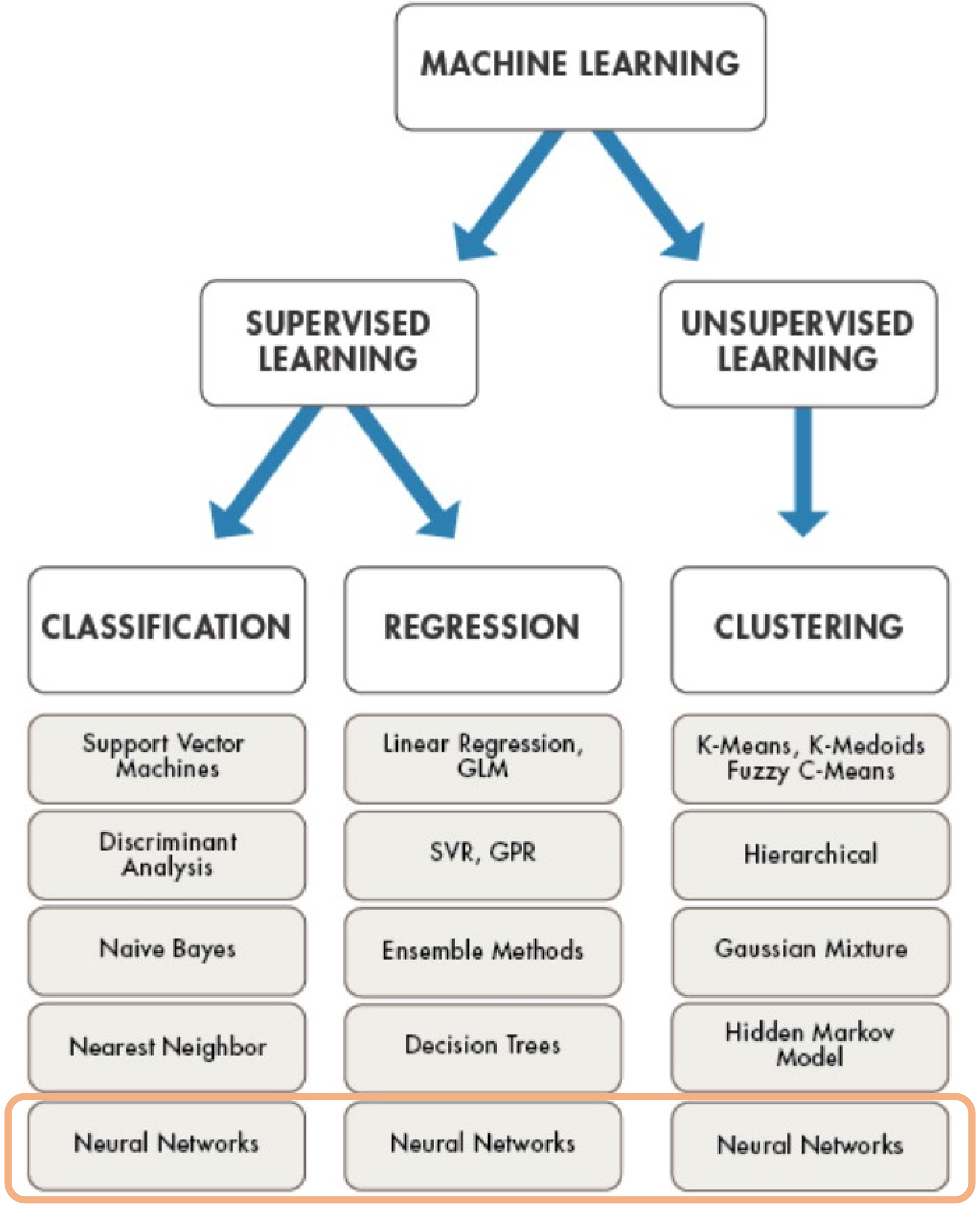
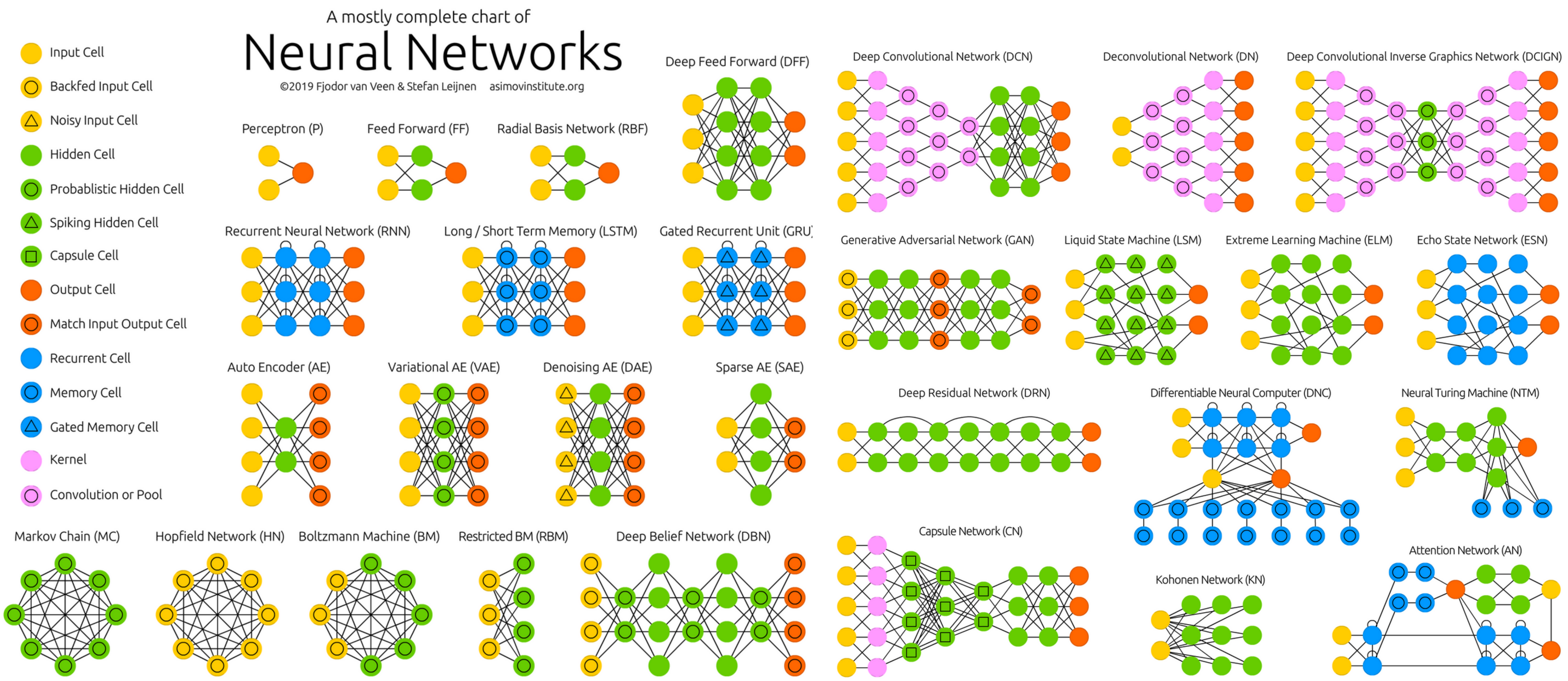


Image from mathworks.com



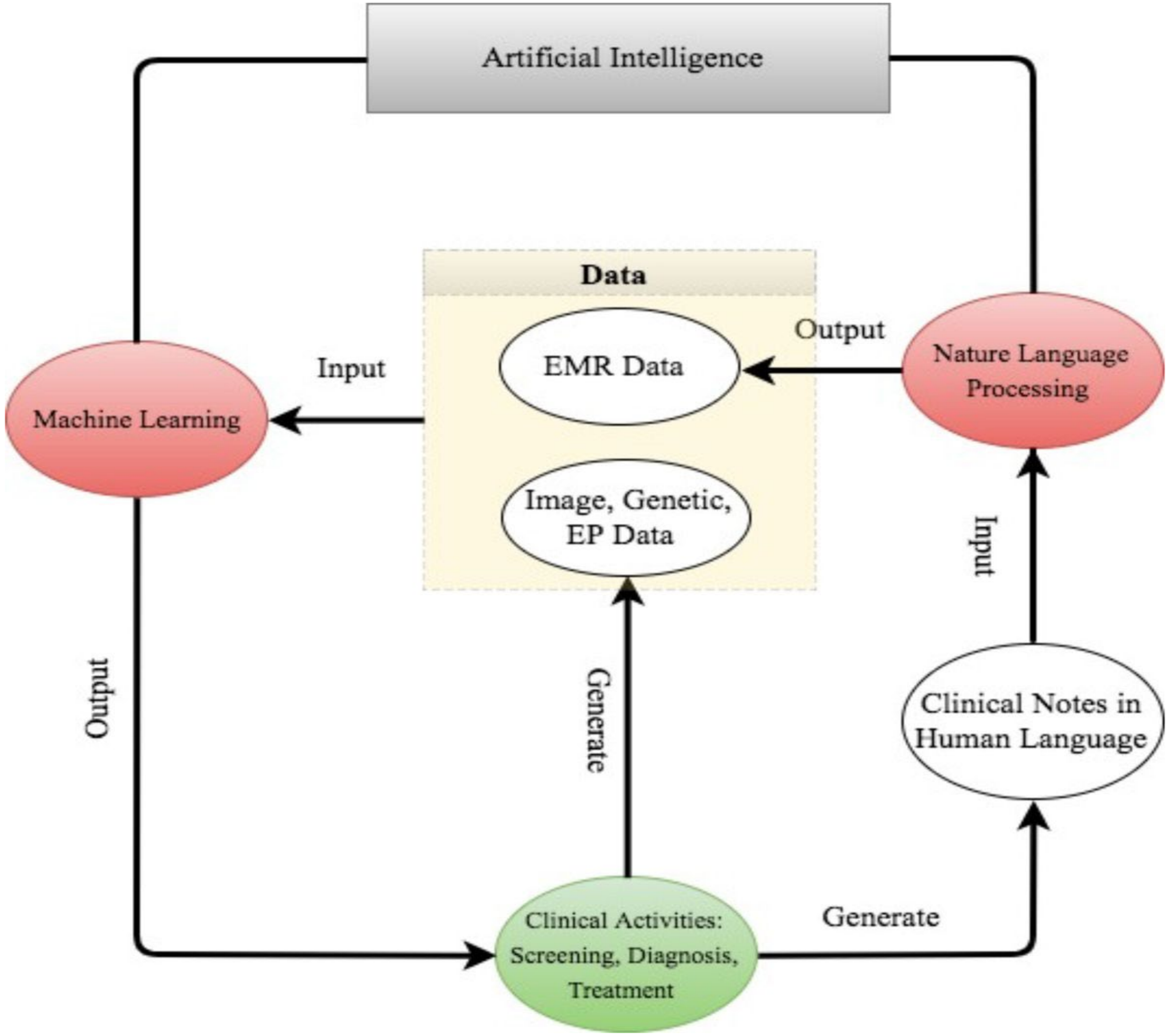
From https://www.asimovinstitute.org/neural-network-zoo/

Types of AI relevant to health (1)

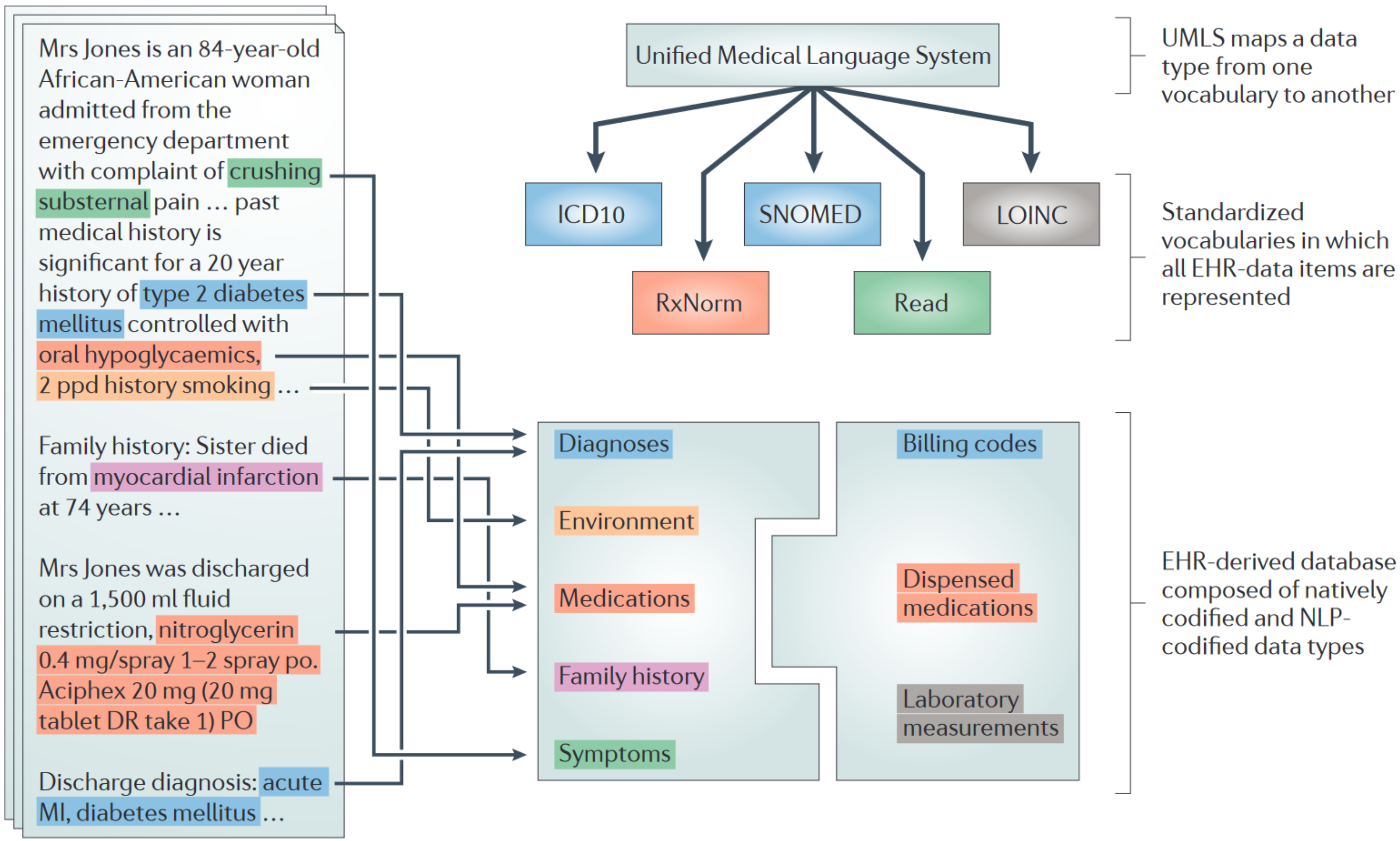
- Machine learning – neural networks and deep learning
  - Precision medicine
  - Predicting what treatment protocols are likely to succeed
  - Prognosticate the evolution of a pathology
  - Recognition of potentially lesions in images

- Natural language processing
  - Statistical NLP (based on deep learning) *vs* Semantic NLP
  - Creation, understanding and classification of clinical documents
  - Conversational AI
  - 80% of free text in medical records

NLP turning texts to machine-readable structured data, which can then be analysed by ML techniques



From https://synthesis-solutions.com/artificial_intelligence_in_healthcare.html



Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011 Jun;12(6):417-28.

Figure 1 | **From clinical notes to structured phenotypes.** Natural language processing (NLP) identifies various concept types in the textual records that are associated with each patient for each medical record.

Types of AI relevant to health (2)

- Rule-based expert systems

Human-Readable Format

```
IF     the stain of the organism is gram negative
AND    the morphology of the organism is rod
AND    the aerobiocity of the organism is gram anaerobic
THEN   there is strong evidence (0.8)
       that the class of the organism is enterobacteriaceae
```

MYCIN Format

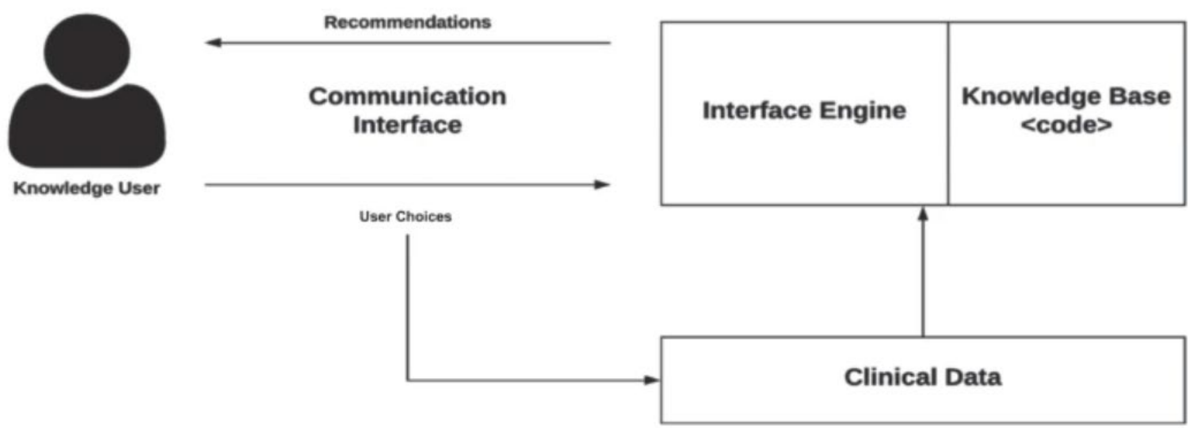
```
IF    (AND  (SAME  CNTEXT  GRAM  GRAMNEG)
            (SAME  CNTEXT  MORPH  ROD)
            (SAME  CNTEXT  AIR  AEROBIC)
THEN  (CONCLUDE  CNTEXT  CLASS  ENTEROBACTERIACEAE
       TALLY  .8)
```

From Seyed Hashem Davarpanah Davarpanah@usc.ac.ir University of Science and Culture. Course Material.

```
data:
/* read the diastolic blood pressure */
diastolic_blood_pressure := read last
{diastolic blood pressure}; /* the value in braces is specific to your
runtime environment */
/* If the height is lower than height_threshold, output a message */
diastolic_pressure_threshold := 60;
stdout_dest := destination
{stdout};
;;
evoke: null_event;;
logic:
if (diastolic_blood_pressure is not number) then
conclude false;
endif;
if (diastolic_blood_pressure >= diastolic_pressure_threshold) then
conclude true;
else
conclude false;
endif;
;;
action:
write "Your Diastolic Blood Pressure is too low (hypotension)"
```

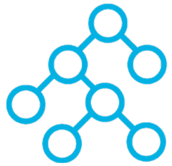
From Wikipedia. https://en.wikipedia.org/wiki/Arden_syntax

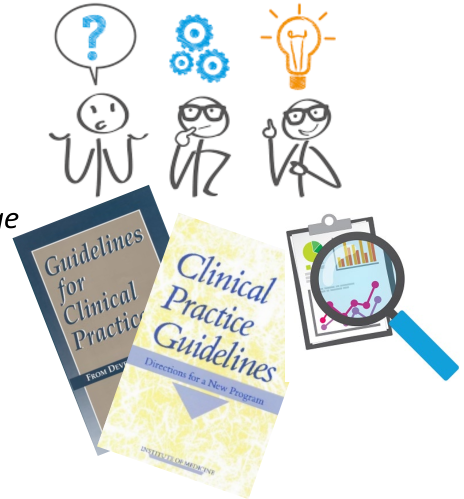
Knowledge based single system CDSS



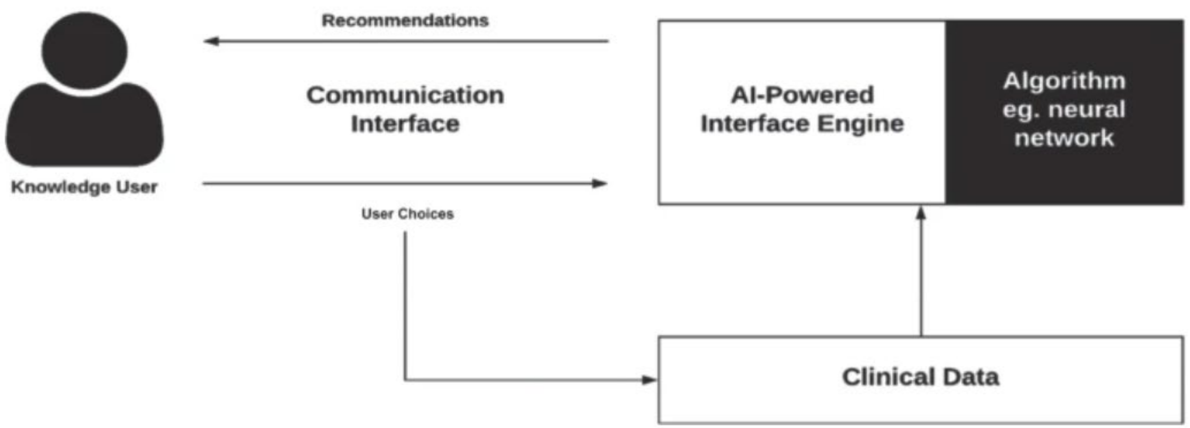
« White box »

Explicit and explicable knowledge

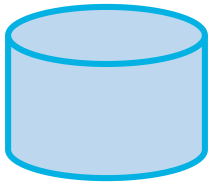




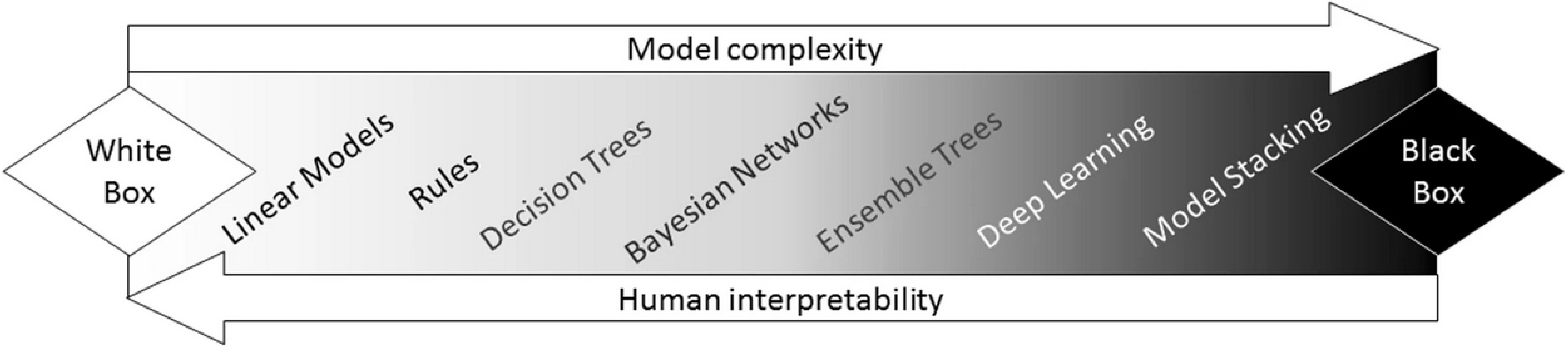
Non-knowledge based single system CDSS



« Black box »



Adapted from Sutton, R.T., Pincock, D., Baumgart, D.C. *et al.* An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digit. Med.* **3,** 17 (2020). https://doi-org.proxy.insermbiblio.inist.fr/10.1038/s41746-020-0221-y



Machine learning models: white- and black-boxes. Increasing model complexity can lead to better approximation of functions and enhance prediction performance, but can lead to a decrease in interpretability of the model

From : Prosperi, M., Min, J. S., Bian, J., & Modave, F. (2018). Big data hurdles in precision medicine and precision public health. *BMC Medical Informatics and Decision Making*, *18*(1), 139. https://doi.org/10.1186/s12911-018-0719-2

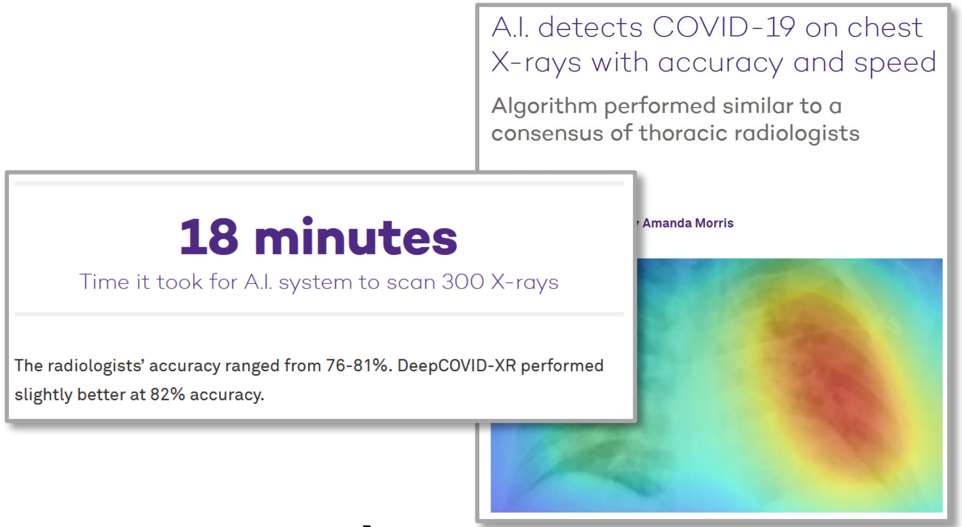
Potential for AI in healthcare and public health

- Diagnosis and treatment applications
  - Good performance (compare to individual expert) but in well-defined and limited areas
  - Need for better integration with clinician workflows and HIS
- Patient engagement and adherence applications
  - Wearable devices
  - Take into account the patient's patterns and self-data
  - To influence the patient's behaviour (nudge)
- Administrative applications, management and planning
  - Claim processing, clinical documentation, medical record management
  - EMR design/usability improvement, chatbots
  - Identifying and eliminating fraud or waste, scheduling patients

Potential for AI in healthcare and public health

- Expanding access to care in underserved or developing regions
  - AI to mitigate the deficit of qualified staff
  - AI to help during care overload

- Diseases prevention
  - Primary and secondary prevention
  - EHR as a risk predictor (but data quality and formats issues)
  - Wearable and personal devices



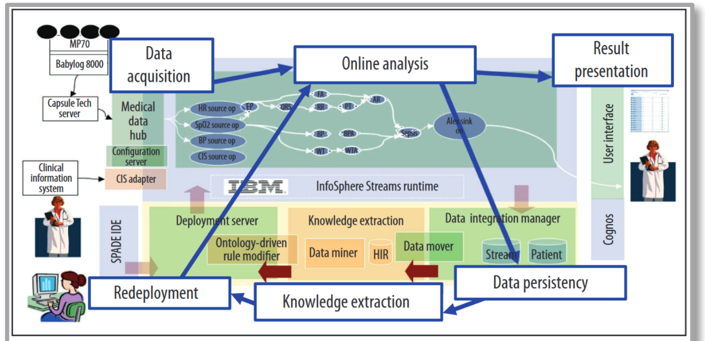
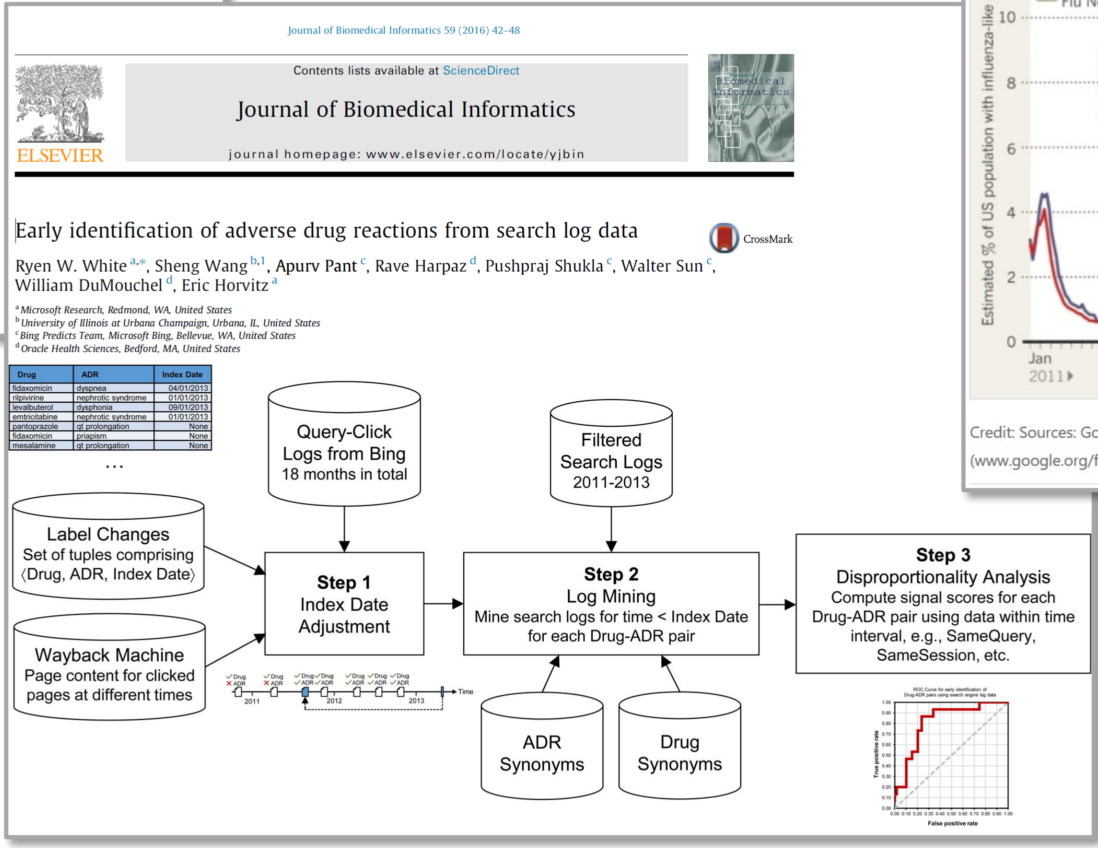
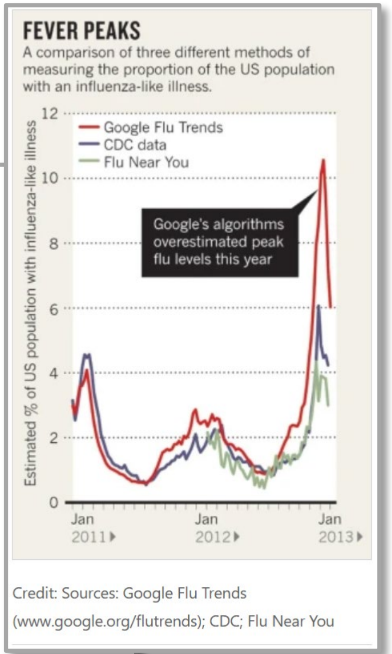


Figure 2. Artemis platform. Artemis enables concurrent diagnoses of multiple patients through real-time analysis of multiple data streams.

Potential for AI in healthcare and public health

- Disease outbreaks and support surveillance

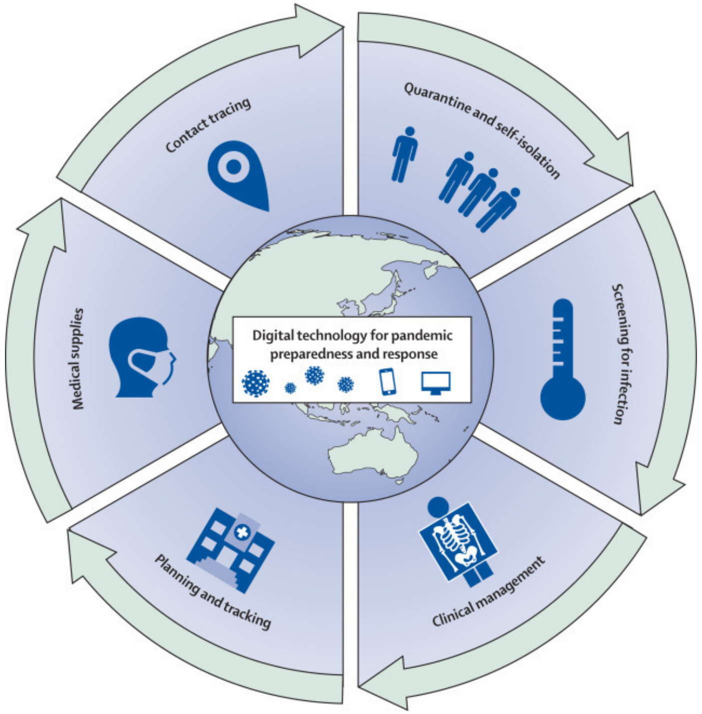




For further examples see
https://www.who.int/publications/i/item/9789240029200

Potential for AI in healthcare and public health

- Disease outbreaks and support surveillance: covid-19 crisis



	Functions	Digital technology	Countries	Advantages	Disadvantages
Tracking	Tracks disease activity in real time	Data dashboards; migration maps; machine learning; real-time data from smartphones and wearable technology	China; Singapore; Sweden; Taiwan; USA	Allows visual depiction of spread; directs border restrictions; guides resource allocation; informs forecasts	Could breach privacy; involves high costs; requires management and regulation
Screening for infection	Screens individuals and populations for disease	Artificial intelligence; digital thermometers; mobile phone applications; thermal cameras; web-based toolkits	China; Iceland; Singapore; Taiwan	Provides information on disease prevalence and pathology; identifies individuals for testing, contact tracing, and isolation	Could breach privacy; fails to detect asymptomatic individuals if based on self-reported symptoms or monitoring of vital signs; involves high costs; requires management and regulation; requires validation of screening tools
Contact tracing	Identifies and tracks individuals who might have come into contact with an infected person	Global positioning systems; mobile phone applications; real-time monitoring of mobile devices; wearable technology	Germany; Singapore; South Korea	Identifies exposed individuals for testing and quarantine; tracks viral spread	Could breach privacy; might detect individuals who have not been exposed but have had contact; could fail to detect individuals who are exposed if the application is deactivated, the mobile device is absent, or Wi-Fi or cell connectivity is inadequate
Quarantine and self-isolation	Identifies and tracks infected individuals, and implements quarantine	Artificial intelligence; cameras and digital recorders; global positioning systems; mobile phone applications; quick response codes	Australia; China; Iceland; South Korea; Taiwan	Isolates infections; restricts travel	Violates civil liberties; could restrict access to food and essential services; fails to detect individuals who leave quarantine without devices
Clinical management	Diagnoses infected individuals; monitors clinical status; predicts clinical outcomes; provides capacity for telemedicine services and virtual care	Artificial intelligence for diagnostics; machine learning; virtual care or telemedicine platforms	Australia; Canada; China; Ireland; USA	Assists with clinical decision-making, diagnostics, and risk prediction; enables efficient service delivery; facilitates patient-centred, remote care; facilitates infection control	Could breach privacy; fails to accurately diagnose patients; involves high costs; equipment may malfunction

Table: **Digital technology initiatives used in pandemic preparedness and response**

From Whitelaw, Sera, et al. « Applications of Digital Technology in COVID-19 Pandemic Planning and Response ». *The Lancet Digital Health*. https://doi.org/10.1016/S2589-7500(20)30142-4.

Potential for AI in healthcare and public health

- Reducing costs (but lack of economic impact assessment studies)

How artificial intelligence can be used to reduce costs and improve outcomes in total joint replacement surgery

Use of AI led to cost reductions including a 25 percent drop in hospital length of stay and 91 percent reduction in discharges to nursing facilities.

Jeff Lagasse, *Associate Editor*

Original Article

The role of prehabilitation with a telerehabilitation system prior to total knee arthroplasty

Morad Chughtai[1], Neil V. Shah[2], Assem A. Sultan[1], Maximillian Solow[3], John V. Tiberi[4], Nima Mehran[5], Trevor North[6], Joseph T. Moskal[7], Jared M. Newman[2], Linsen T. Samuel[1], Anil Bhave[8], Michael A. Mont[1,9]

JOURNAL OF MEDICAL INTERNET RESEARCH Wolff et al

Review

The Economic Impact of Artificial Intelligence in Health Care: Systematic Review

Justus Wolff[1,2*], MSc; Josch Pauling[1*], PhD; Andreas Keck[2*], MD; Jan Baumbach[1*], PhD

[1]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[2]Strategy Institute for Digital Health, Hamburg, Germany
[*]all authors contributed equally

Potential for AI in healthcare and public health

Use Case or User Group	Category	Examples of Applications	Technology
Patients and families	Health monitoring Benefit/risk assessment	• Devices and wearables • Smartphone and tablet apps, websites	Machine learning, natural language processing (NLP), speech recognition, chatbots
	Disease prevention and management	• Obesity reduction • Diabetes prevention and management • Emotional and mental health support	Conversational AI, NLP, speech recognition, chatbots
	Medication management	• Medication adherence	Robotic home telehealth
	Rehabilitation	• Stroke rehabilitation using apps and robots	Robotics
Clinician care teams	Early detection, prediction, and diagnostics tools	• Imaging for cardiac arrhythmia detection, retinopathy • Early cancer detection (e.g., melanoma)	Machine Learning
	Surgical proce-dures	• Remote-controlled robotic surgery • AI-supported surgical roadmaps	Robotics, machine learning
	Precision medicine	• Personalized chemotherapy treatment	Supervised machine learning, reinforcement learning
	Patient safety	• Early detection of sepsis	Machine learning

Use Case or User Group	Category	Examples of Applications	Technology
Public health program managers	Identification of individuals at risk	• Suicide risk identification using social media	Deep learning (convolutional and recurrent neural networks)
	Population health	• Eldercare monitoring	Ambient AI sensors
	Population health	• Air pollution epidemiology • Water microbe detection	Deep learning, geospatial pattern mining, machine learning
Business administrators	International Classification of Diseases, 10th Rev. (ICD-10) coding	• Automatic coding of medical records for reimbursement	Machine learning, NLP
Business administrators	Fraud detection	• Health care billing fraud • Detection of unlicensed providers	Supervised, unsupervised, and hybrid machine learning
	Cybersecurity	• Protection of personal health information	Machine learning, NLP
	Physician management	• Assessment of physician competence	Machine learning, NLP
Researchers	Genomics	• Analysis of tumor genomics	Integrated cognitive computing
	Disease prediction	• Prediction of ovarian cancer	Neural networks
	Discovery	• Drug discovery and design	Machine learning, computer-assisted synthesis

From "*Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*". National Academy of Medicine.
https://nam.edu/artificial-intelligence-special-publication/

Challenges

Challenge	Description
Workflow integration	Understand the technical, cognitive, social, and political factors in play and incentives impacting integration of AI into health care workflows.
Enhanced explainability and interpretability	To promote integration of AI into health care workflows, consider what needs to be explained and approaches for ensuring understanding by all members of the health care team.
Workforce education	Promote educational programs to inform clinicians about AI/machine learning approaches and to develop an adequate workforce.
Oversight and regulation	Consider the appropriate regulatory mechanism for AI/machine learning and approaches for evaluating algorithms and their impact.
Problem identification and prioritization	Catalog the different areas of health care and public health where AI/machine learning could make a difference, focusing on intervention-driven AI.
Clinician and patient engagement	Understand the appropriate approaches for involving consumers and clinicians in AI/machine learning prioritization, development, and integration, and the potential impact of AI/machine learning algorithms on the patient-provider relationship.
Data quality and access	Promoting data quality, access, and sharing, as well as the use of both structured and unstructured data and the integration of non-clinical data is critical to developing effective AI tools.

From "*Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*". National Academy of Medicine.
https://nam.edu/artificial-intelligence-special-publication/

SESTIM

Challenges

- Implement Learning Health Systems (LHS)

  “health system in which internal data and experience are systematically integrated with external evidence, and that knowledge is put into practice”

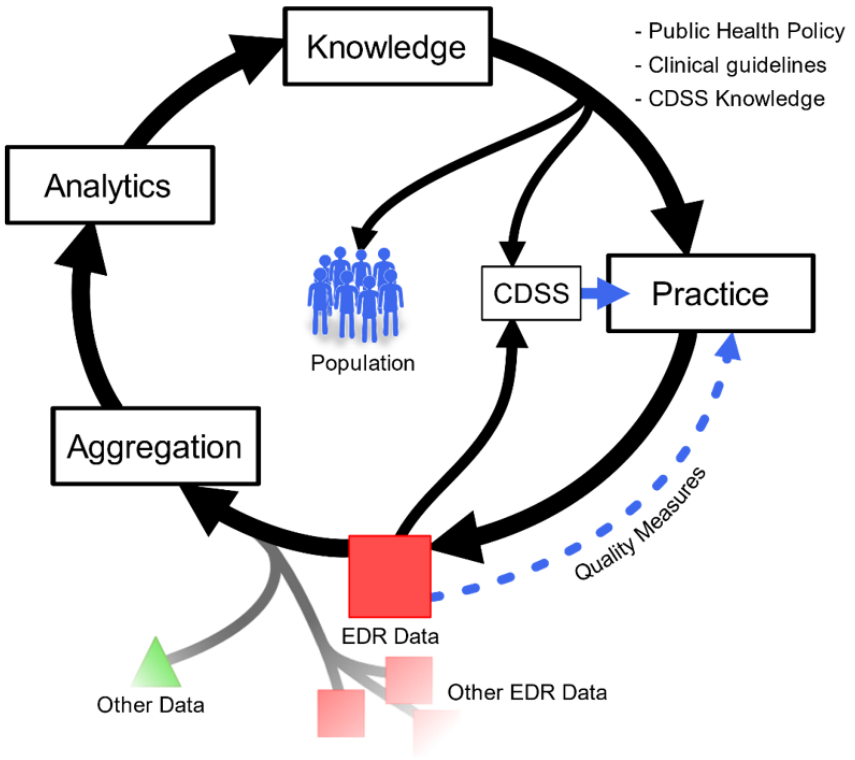
  (AHRQ definition)

Challenges

- Implement Learning Health Systems (LHS)

  “health system in which internal data and experience are systematically integrated with external evidence, and that knowledge is put into practice”

  (AHRQ definition)



Challenges : LHS and AI in health

Topic	IOM Learning Health System Recommendation	Mapping to AI in Health Care
Foundational Elements		
Digital infrastructure	Improve the capacity to capture clinical, care delivery process, and financial data for better care, system improvement, and the generation of new knowledge.	Improve the capacity for unbiased, representative data capture with broad coverage for data elements needed to train AI.
Data utility	Streamline and revise research regulations to improve care, promote and capture clinical data, and generate knowledge.	Leverage continuous quality improvement (QI) and implement scientific methods to help select when AI tools are the most appropriate choice to optimize clinical operations and harness AI tools to support continuous improvement.
Care Improvement Targets		
Clinical decision support	Accelerate integration of the best clinical knowledge into care decisions.	Accelerate integration of AI tools into clinical decision support applications.
Patient-centered care	Involve patients and families in decisions regarding health and health care, tailored to fit their preferences.	Involve patient and families in how, when, and where AI tools are used to support care in alignment with preferences.
Community links	Promote community–clinical partnerships and services aimed at managing and improving health at the community level.	Promote use of AI tools in community and patient health consumer applications in a responsible, safe manner.
Care continuity	Improve coordination and communication within and across organizations.	Improve AI data inputs and outputs through improved card coordination and data interchange.
Optimized operations	Continuously improve health care operations to reduce waste, streamline care delivery, and focus on activities that improve patient health.	Leverage continuous QI and Implementation Science methods to help select when AI tools are the most appropriate choice to optimize clinical operations.

Topic	IOM Learning Health System Recommendation	Mapping to AI in Health Care
Policy Environment		
Financial incentives	Structure payment to reward continuous learning and improvement in the provision of best care at lower cost.	Use AI tools in business practices to optimize reimbursement, reduce cost, and (it is hoped) do so at a neutral or positive balance on quality of care.
Performance transparency	Increase transparency on health care system performance.	Make robust performance characteristics for AI tools transparent and assess them in the populations within which they are deployed.
Broad leadership	Expand commitment to the goals of a continuously learning health care system.	Promote broad stakeholder engagement and ownership in governance of AI systems in health care.

From "*Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*". National Academy of Medicine. https://nam.edu/artificial-intelligence-special-publication/

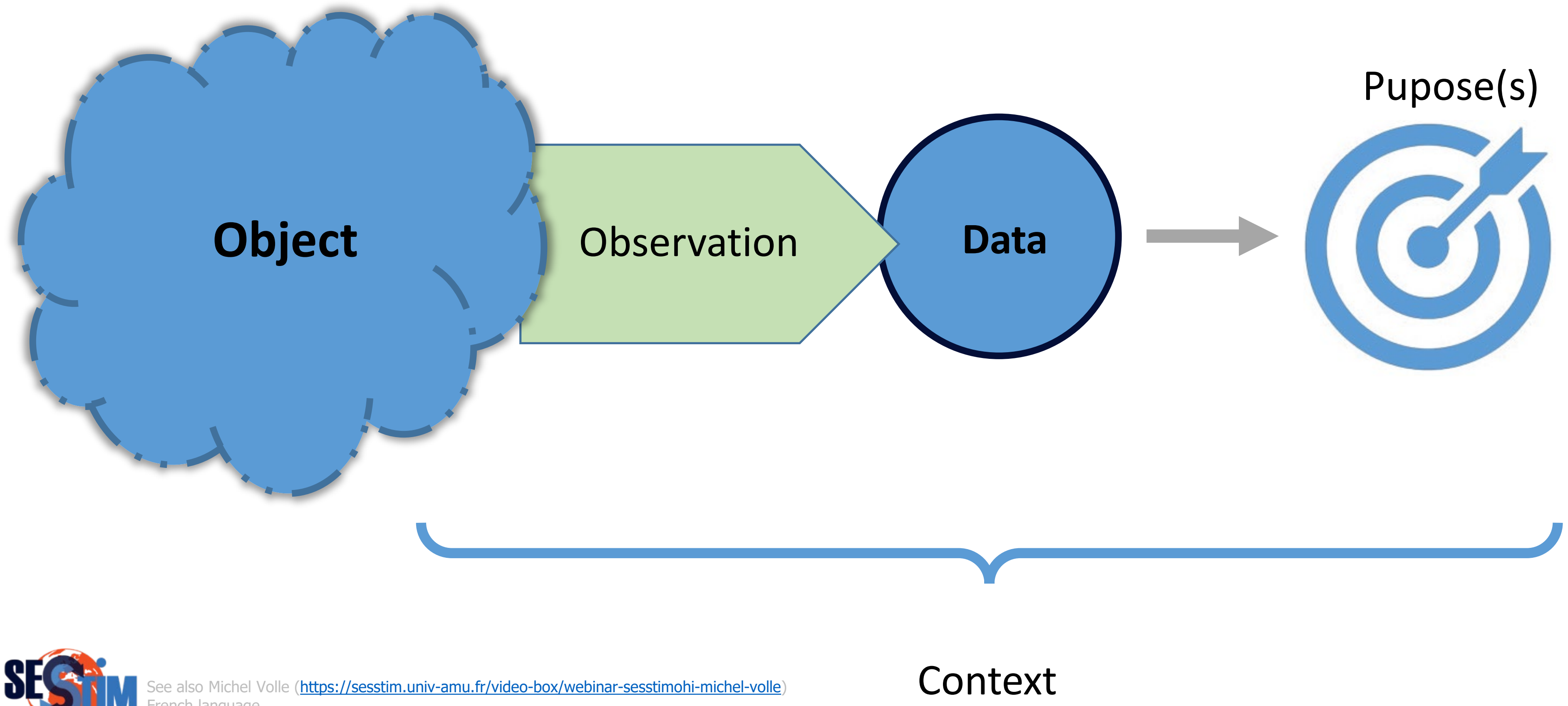
Data for AI

- Data are critical for developing AI algorithms
- Key technologies for data capture are now well established (smartphone, mobile apps and device, EMR, HIS,…)
  - → Heterogeneous data-rich environment
- Data are mainly collected via information systems
  - → Knowledge of the underlying processes and IS specificities is mandatory
- Data are collected with an initial/primary objective
  - → Reusing data (for AI analyzes) must takes into account this objective

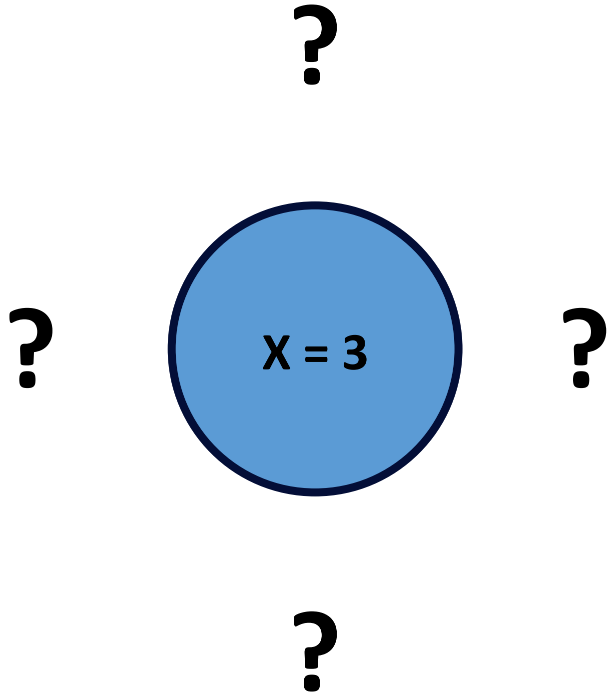
SESTIM

What are data, information, knowledge…and wisdom ?

Data are selective observations

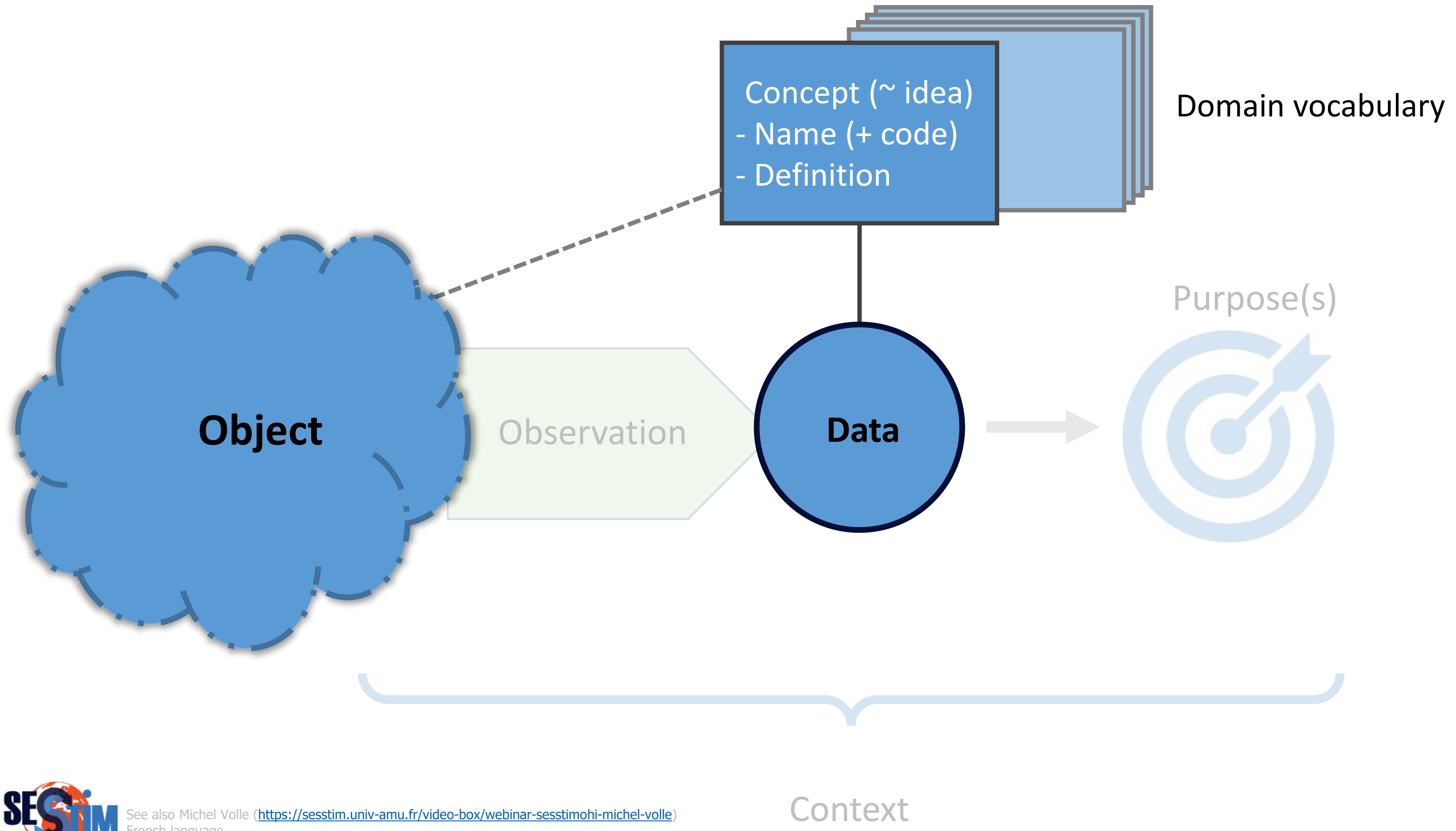


Isolated data are meaningless



See also Michel Volle (https://sesstim.univ-amu.fr/video-box/webinar-sesstimohi-michel-volle)
French language

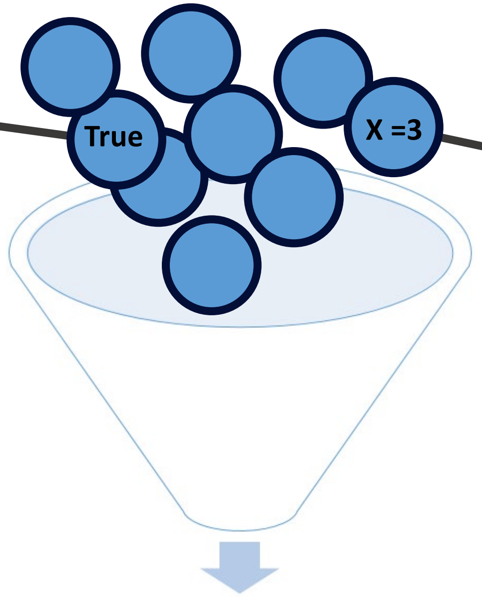
Data are linked to concepts



Data + knowledge => relevant action

- Knowledge : relationships/laws/general rules applicable to a data set

Concept
-name : Driver
-definition : A person responsible for guiding a vehicle..



Concept
-name : Alcohol level
-unit : gr/L
-location : blood
-definition : Mass of alcohol per unit volume of BLOOD.

Knowledge :
- Alcohol levels above 0.5 g/L impair reflexes
- Good reflexes are needed to drive

Decision
=
« do not drive! »

See also Michel Volle (https://sesstim.univ-amu.fr/video-box/webinar-sesstimohi-michel-volle)
French language

Data > Information > Knowledge > Wisdom

- ***Data*** *are undigested observations, unvarnished facts.*
- ***Information*** *is organized data—organized by others, not by me.*
- ***Knowledge*** *is organized information, internalized by me, integrated with everything else I know from experience or study or intuition, and therefore useful in guiding my life and work.*
- ***Wisdom*** *is integrated knowledge, information made super useful by theory, which relates bits and fields of knowledge to each other, which in turn enables me to use the information to do something.*

Cleveland H. cited by Baker, E. L., Fond, M., Hale, P., & Cook, J. (2016). What is "informatics"? Journal of Public Health Management and Practice, 22(4), 420-423. https://doi.org/10.1097/PHH.0000000000000415

➡ Health Information Systems play a role in all these elements

Structure of computerized data

- « Simple » structure (attribute: value)

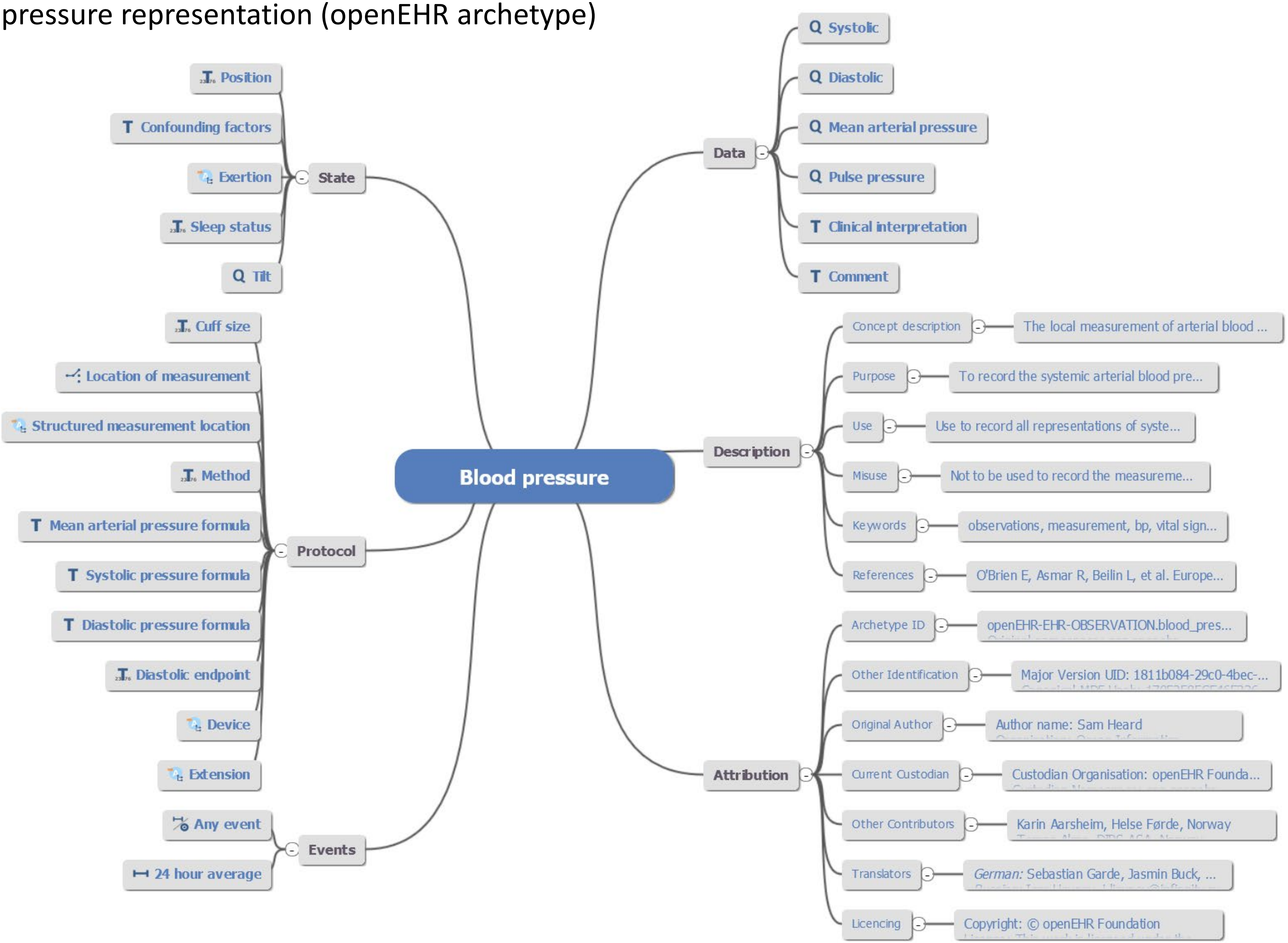
  Examples:
  - Blood alcohol: 3,2 g/L
  - Asthenia: True
  - Eye colour: Blue

- « composite » structure

  Example:
  - Blood pressure
    - Systolic blood pressure: 12 mmHg
    - Diastolic blood pressure: 8 mmHg

Example of blood pressure representation (openEHR archetype)

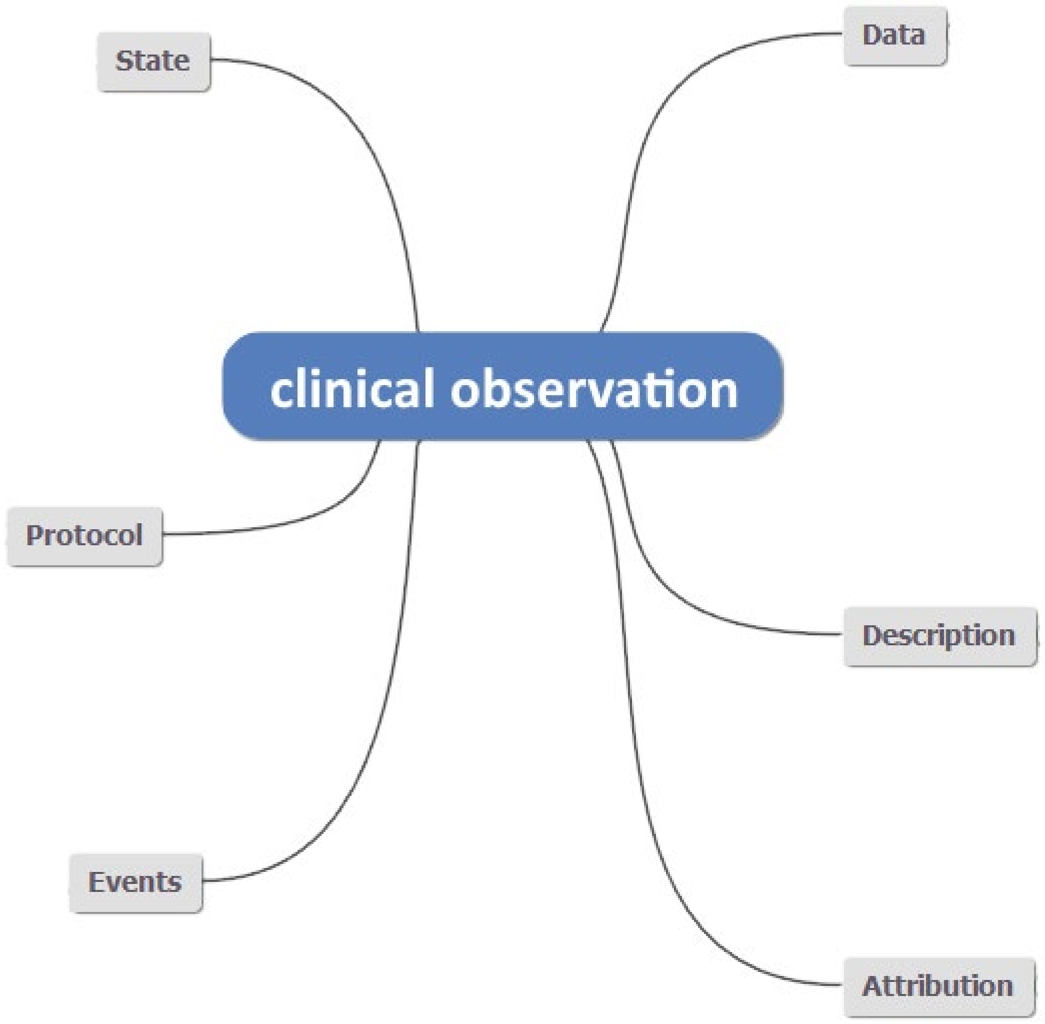
https://ckm.openehr.org



Structure <-> Model

Model

Representation and organization of elements considered as essential for an observed reality, in a given context and for a given purpose(s)



Structure <-> Syntax

Syntax

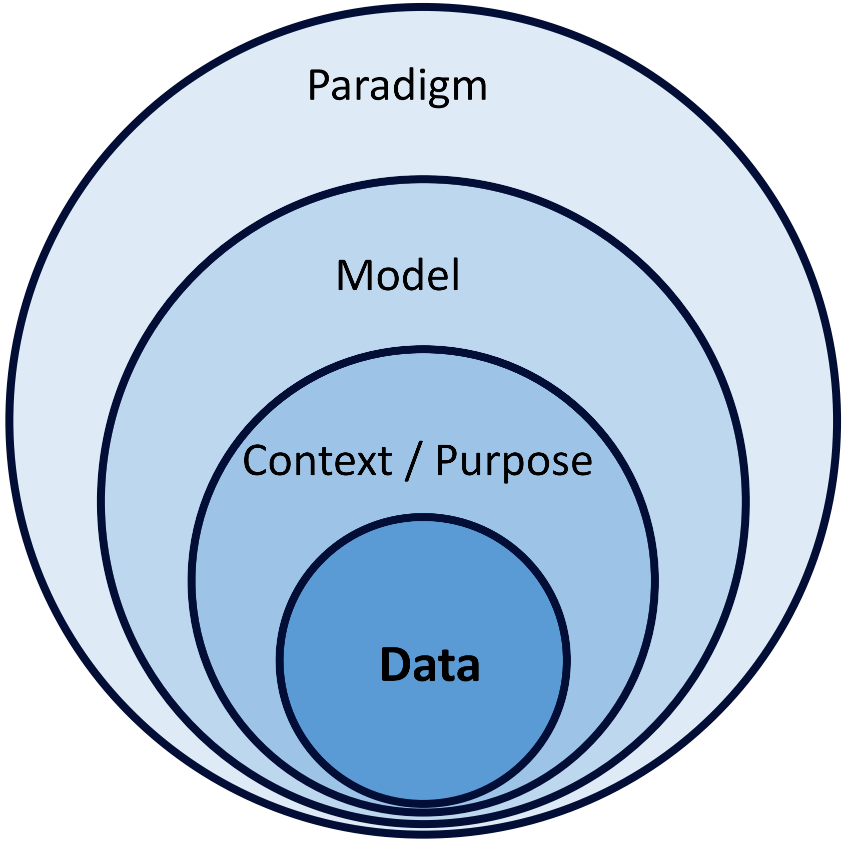
Rules on how to write
and dispose information/data

Example of a JSON syntax

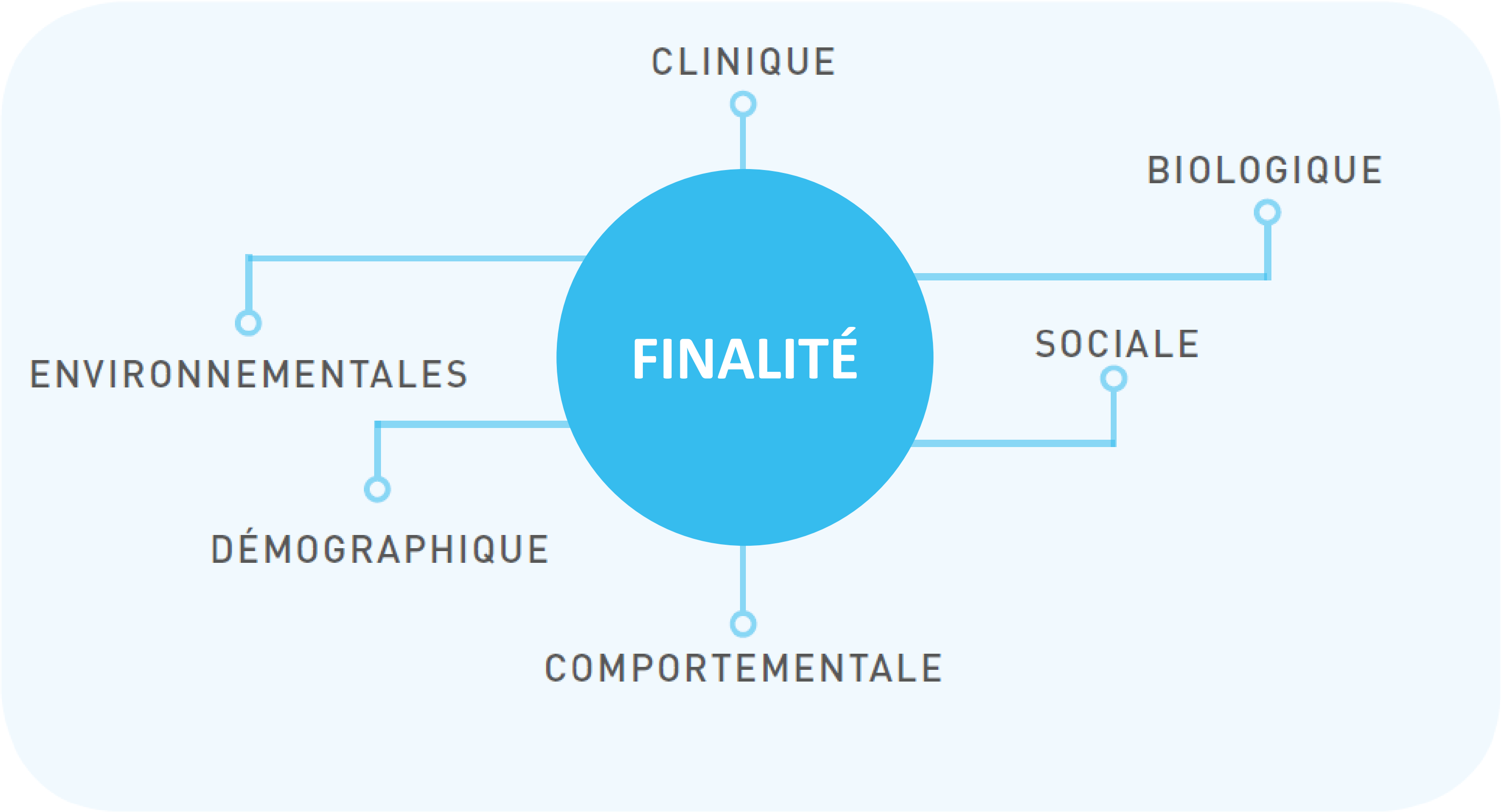
```
 "resourceType": "Observation",
 "id": "blood-pressure",
…
 "bodySite": {
  "coding": [
   {
    "system": "http://snomed.info/sct",
    "code": "368209003",
    "display": "Right arm"
…
 "component": [
  {
   "code": {
    "coding": [
     {
      "system": "http://loinc.org",
      "code": "8480-6",
      "display": "Systolic blood pressure"
…
   "valueQuantity": {
    "value": 107,
    "unit": "mmHg",
    "system": "http://unitsofmeasure.org",
    "code": "mm[Hg]"
   },
…
```

```
"display": "normal"
  {
   "code": {
    "coding": [
     {
      "system": "http://loinc.org",
      "code": "8462-4",
      "display": "Diastolic blood pressure"
…
   "valueQuantity": {
    "value": 60,
    "unit": "mmHg",
    "system": "http://unitsofmeasure.org",
    "code": "mm[Hg]"
   },
   "interpretation": [
    {
     "coding": [
      {
       "system": "http://terminology.hl7",
       "code": "L",
       "display": "low"
 …
```

Data are the result of choices

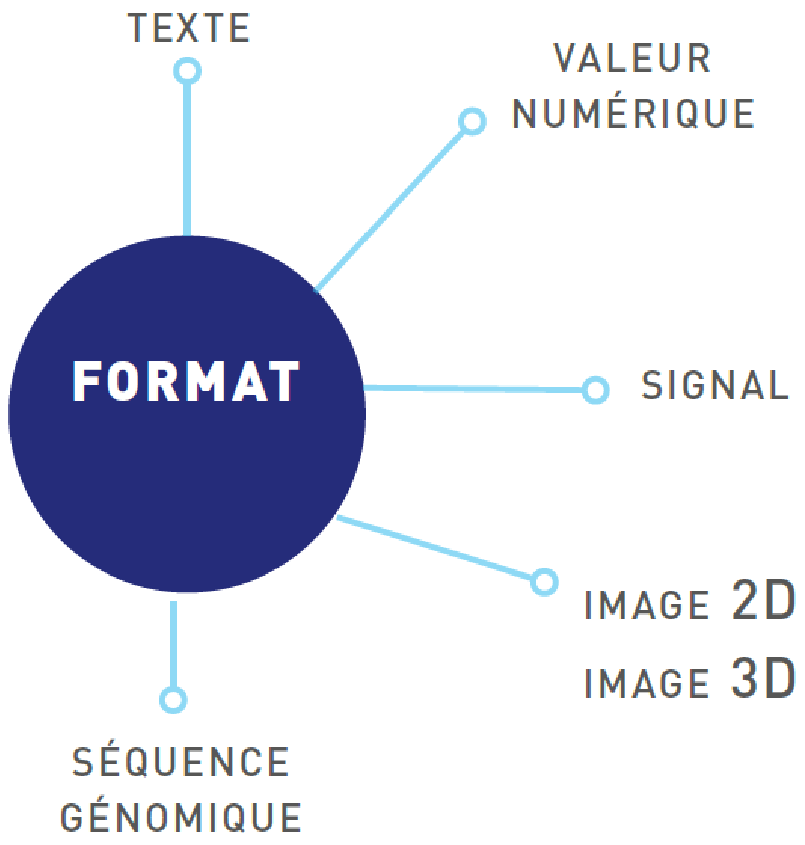


Heterogeneity of health data



Crédits infographie : étude sur l'usage des données de santé (2018) lir.asso.fr

Heterogeneity of health data



Crédits infographie : étude sur l'usage des données de santé (2018) lir.asso.fr

Heterogeneity of health data

SOURCES

ESSAIS CLINIQUES
(PRIVÉS/PUBLICS)

BASES DE DONNÉES
MÉDOCO-ADMINISTRATIVES
SNIIRAM* : 8,9 milliards
de feuilles de soins en 2015,
20 milliards en ligne,
PMSI**

DONNÉES DES PATIENTS
Objets connectés,
applications

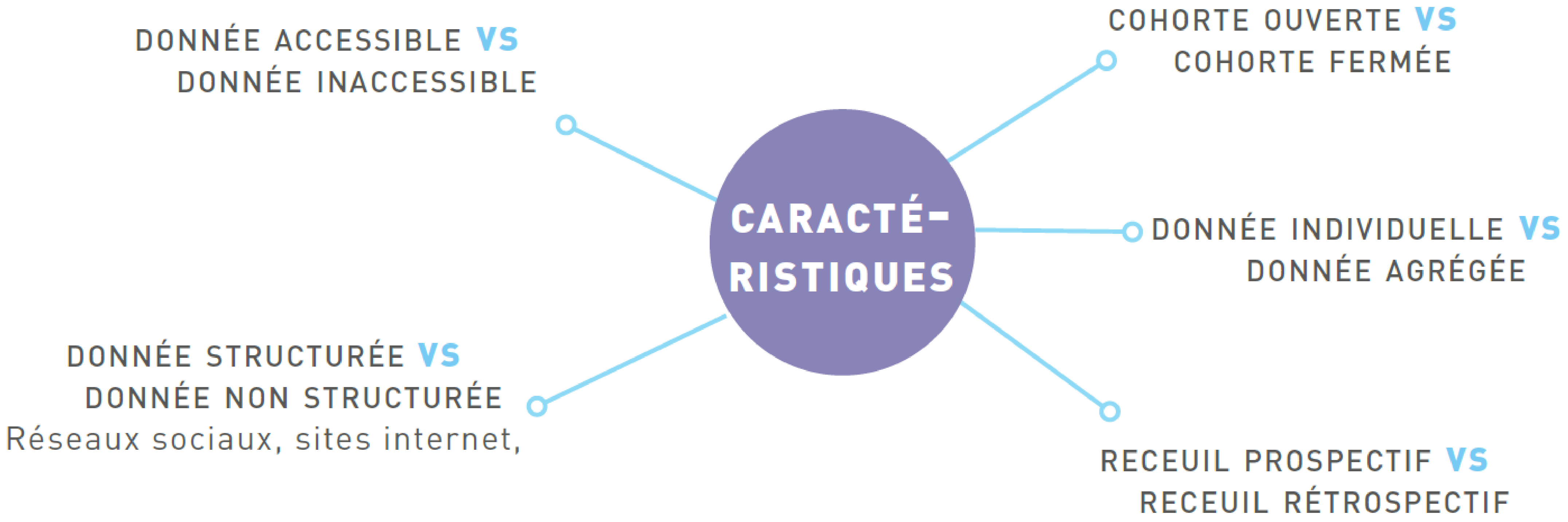
COHORTES/REGISTRES
Constances
(200 000 volontaires
de 18 à 69 ans),
Hepater

DONNÉES D'OPINION
Réseaux sociaux,
sites internet,
moteurs de recherche

DOSSIER MÉDICAL
Etablissements de santé/
professionnels de ville

Crédits infographie : étude sur l'usage des données de santé (2018) lir.asso.fr

Heterogeneity of health data

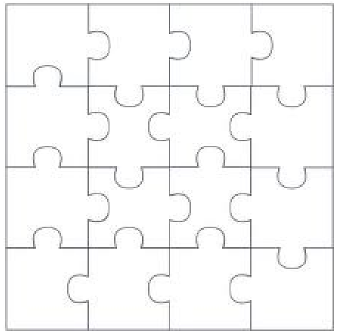


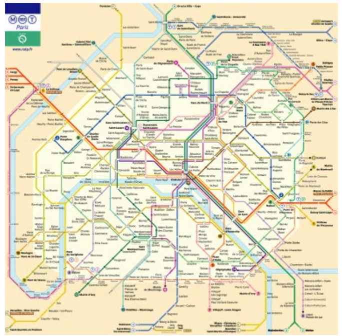
Crédits infographie : étude sur l'usage des données de santé (2018) lir.asso.fr

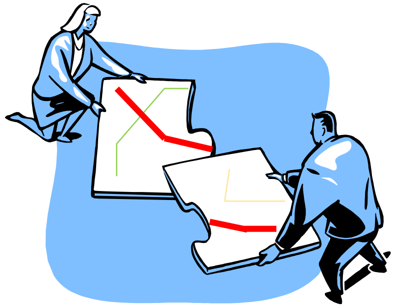
Notion of interoperability

Interoperability of information systems (IS): several levels

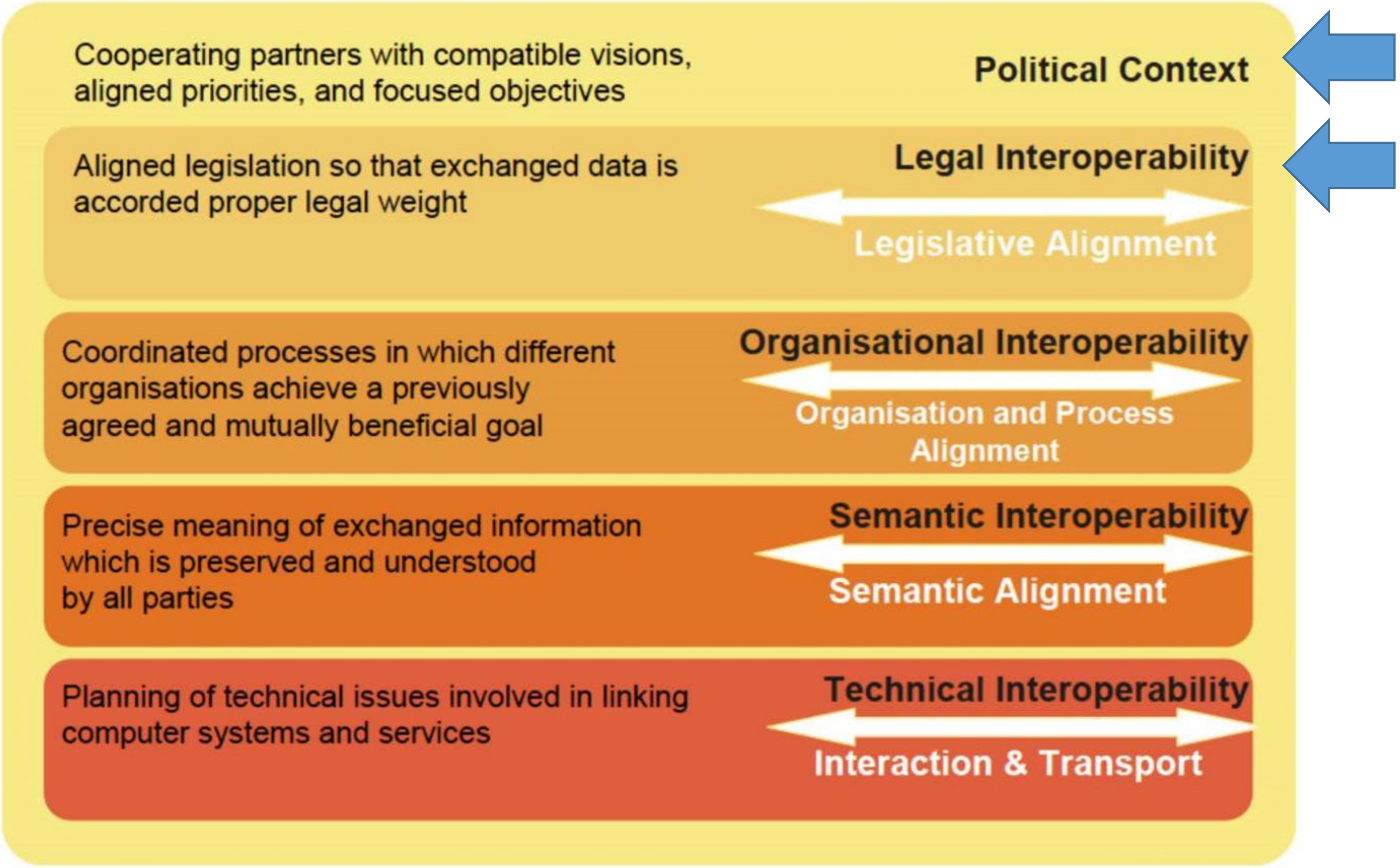
- Syntactic → Technical
- Semantics → Models and concepts (nomenclature/classification/code systems)
- Organisational → Processes alignment







Interoperability levels



See: http://references.modernisation.gouv.fr/interoperabilite anf https://ec.europa.eu/isa2/eif
Video: https://www.youtube.com/watch?v=g-CzHHJ0ZTM&

Developments regarding data and their (re)use

Natively digital data are resources (Big data, AI)

Data-driven research as a complement to traditional hypothesis-driven research

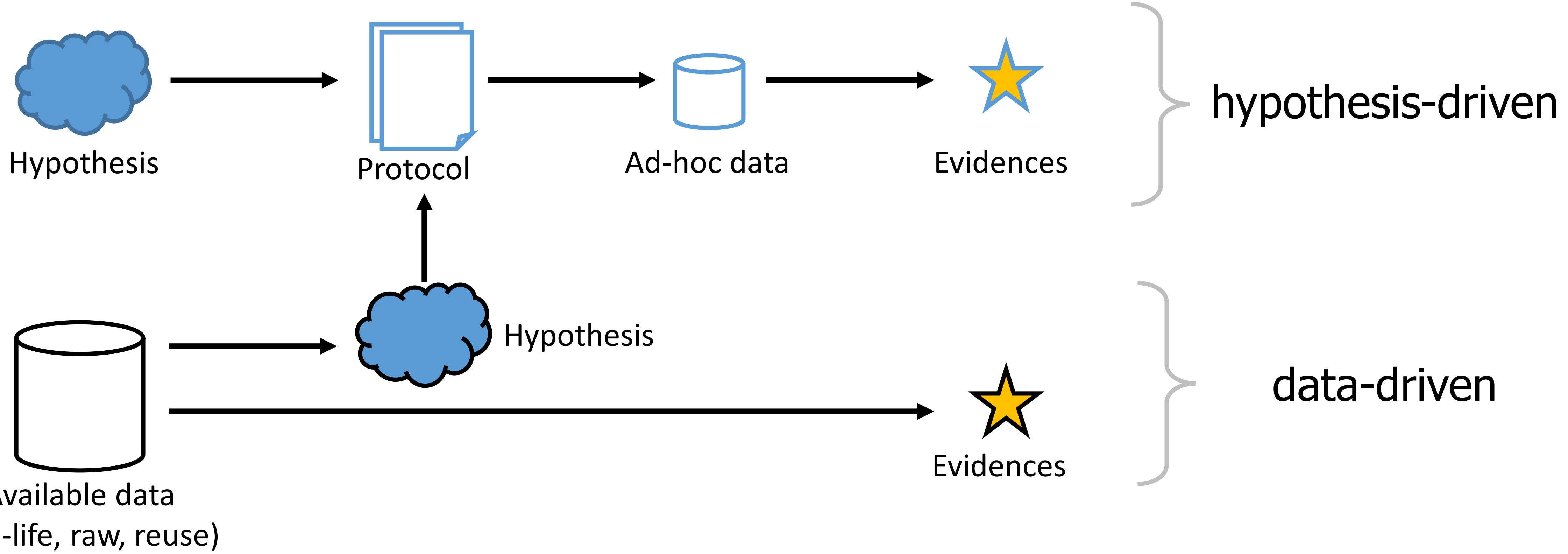


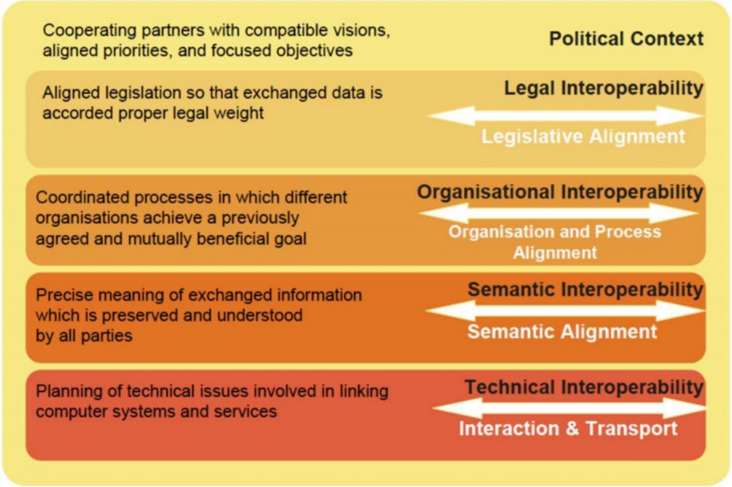
Schéma d'après M. Fieschi

Hypothesis generation

Features selection methods

- Wrapper methods
  - Fitting machine-learning models on different subsets of variables
  - Selection of the subset of variables the most predictive
- Filter methods
  - Studying correlation, p-value, other statistical tests
  - Lower computational costs
- Embedded methods
  - Embed the variable selection step into the learning algorithm
  - Examples: Least Absolute Shrinkage and Selection Operator (LASSO), elastic nets, regularized trees

Barriers hampering data reuse



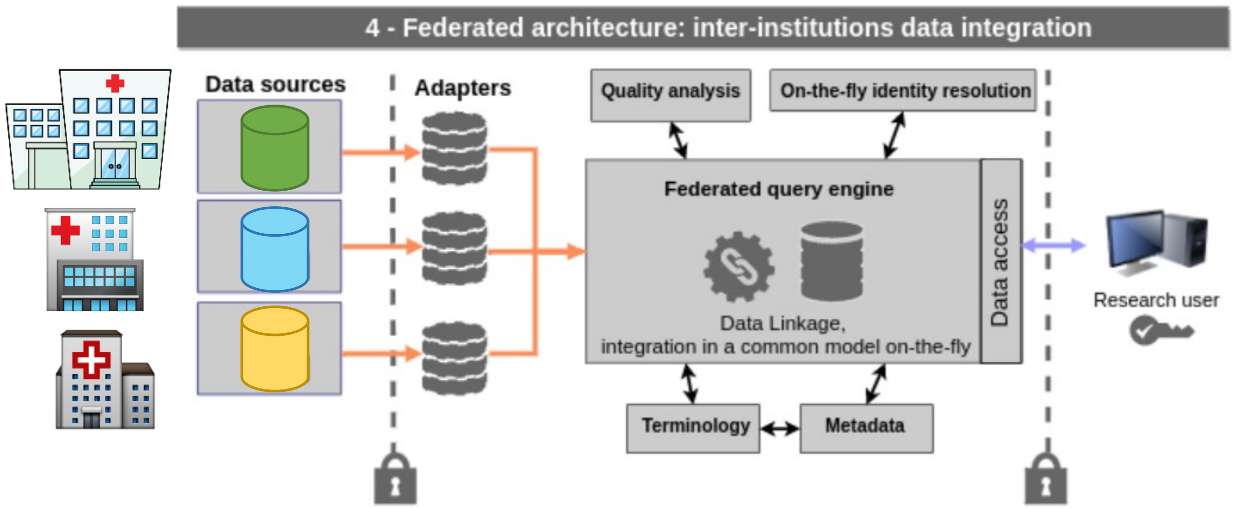
- Heterogeneity of Representations
  - Raw data representations
  - Query languages
  - Communication protocol
  - Vocabulary mismatching
- Knowledge boundaries
  - Explicit knowledge (communicable)
  - Implicit knowledge (not communicable ?)

Thanos, Costantino. « Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies ». *Publications*, vol. 5, nº 1, janvier 2017, p. 2. *DOI.org (Crossref)*, https://doi.org/10.3390/publications5010002.

Barriers hampering data reuse

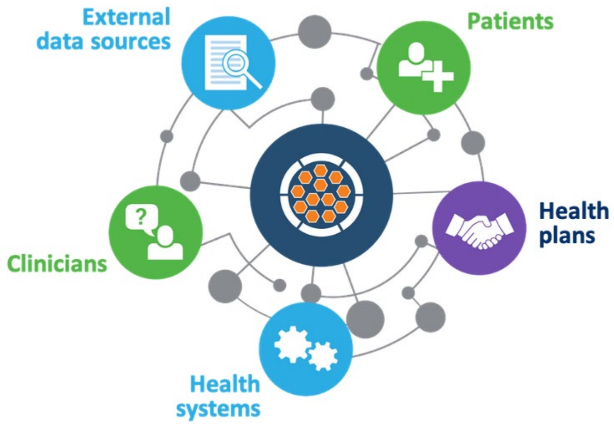
- Discovering data
  - Who has the data needed and where
  - Trust relationships with owner

Inter-instutional research plateform – scientific collaboration



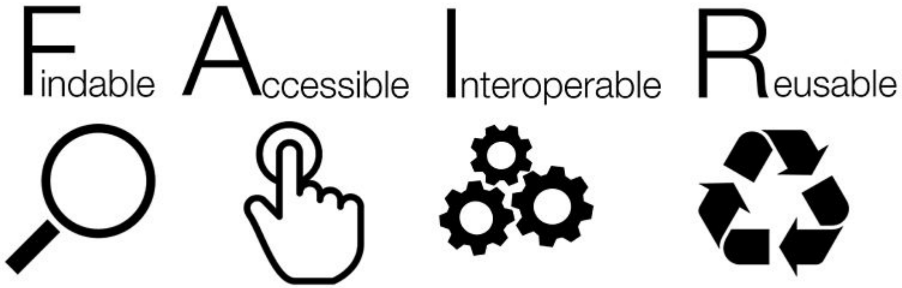
From Gagalova K et al. What You Need to Know Before Implementing a Clinical Research Data Warehouse: Comparative Review of Integrated Data Repositories in Health Care Institutions JMIR Form Res 2020;4(8):e17687

For examples see :
https://www.ohdsi.org - http://www.covidclinical.net - https://pcornet.org - https://i2b2transmart.org/

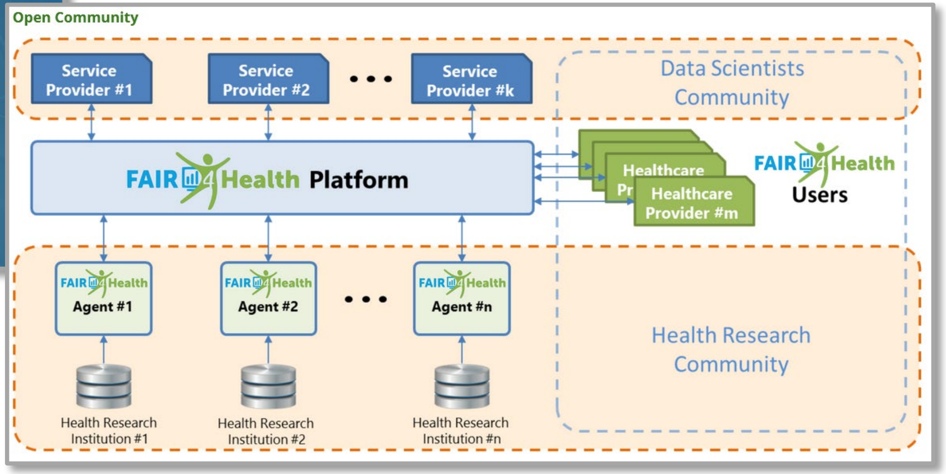


Barriers hampering data reuse

- Discovering data → data publication



https://www.go-fair.org/fair-principles/



https://www.fair4health.eu

Barriers hampering data reuse

Data quality

From Kahn MG, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016 Sep 11;4(1):1244.

1. Conformance (is the <u>format</u> of the data <u>consistent</u> with what is expected ?)
   - Format of the value itself
   - Conformity of the code or label identifying the data
2. Completeness (is the <u>value</u> of the data <u>present</u> ?)
   - Accidental missingness (**MCAR** *Missing Completely at Random*)
   - Context-dependent missingness (**MAR** *Missing At Random*, **MNAR** *Missing Not At Random*)
3. Plausibility
   - Timeliness
   - Temporal

Barriers hampering data reuse

Data bias in AI

- Inappropriate training dataset (under/over-representation)
- Lack of knowledge about data

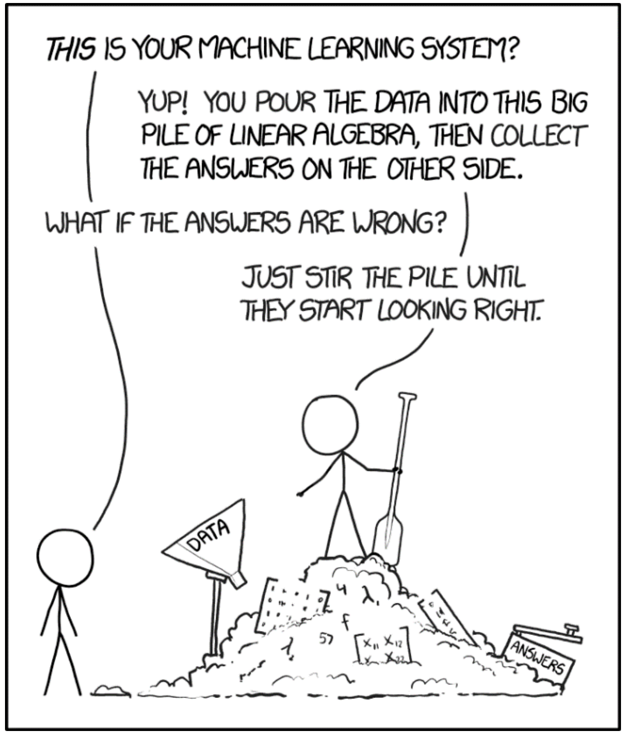
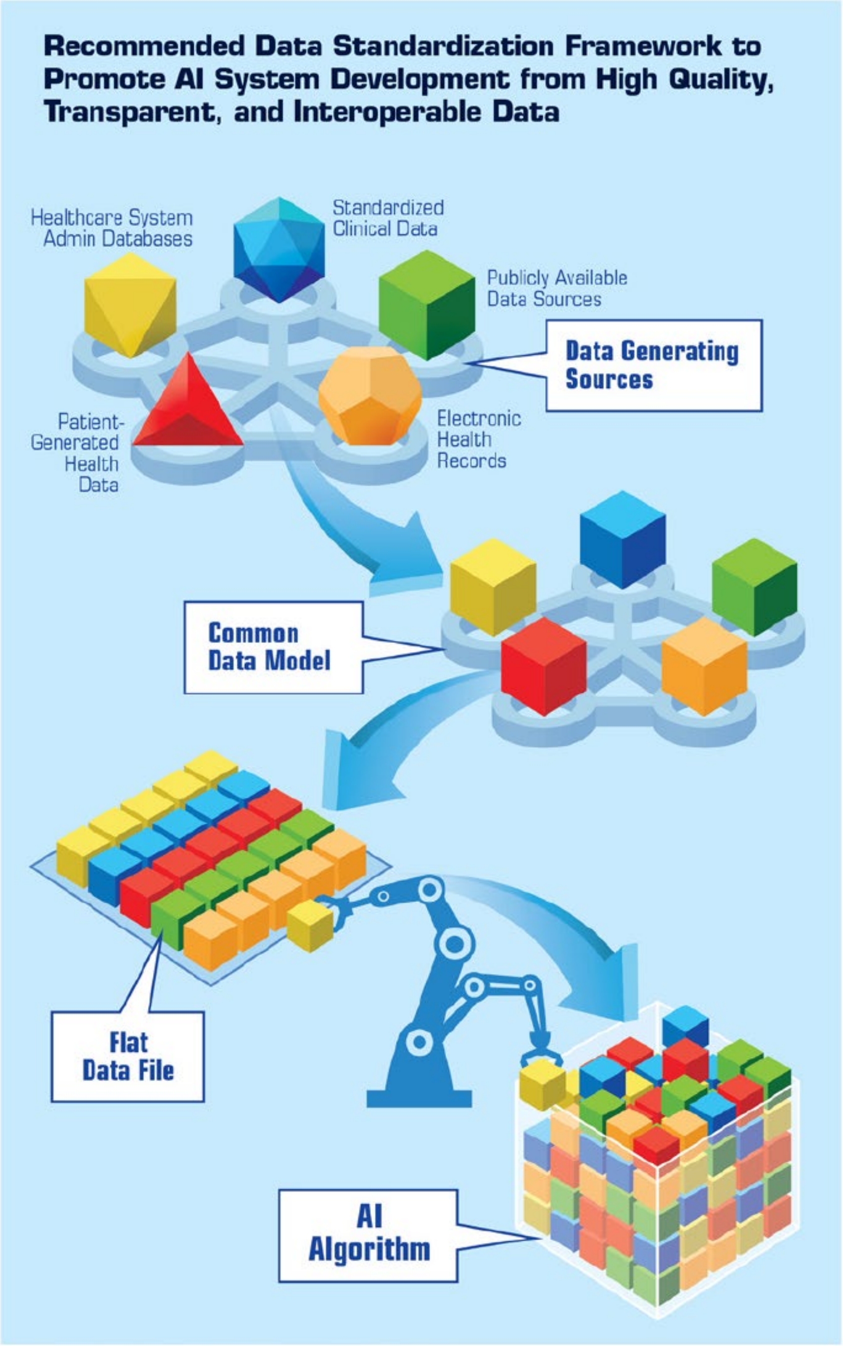


Image from https://towardsdatascience.com

Pathway from raw data to AI algorithm

Heterogeneous data integration is the next challenge for AI

…and an old challenge for HIS !



SESTIM