



3 & 4 octobre 2024

Centre Prouvé - 1 Pl. de la République,
54000 Nancy



14^E CNRC

Intelligence artificielle, définition et principes

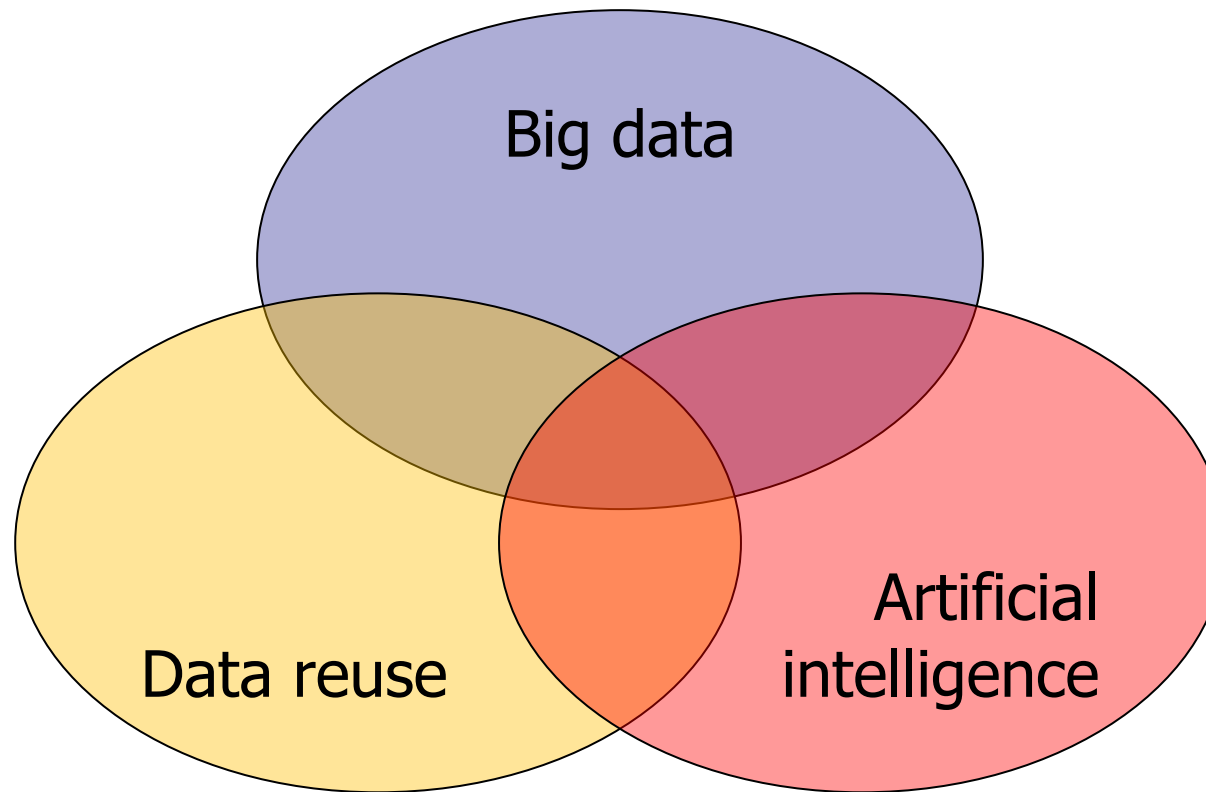
Emmanuel CHAZARD, Lille

Presentation



The ideas expressed in this presentation do not necessarily represent those of the organisms nor of the persons who are cited.

Overlap, frequent confusion...

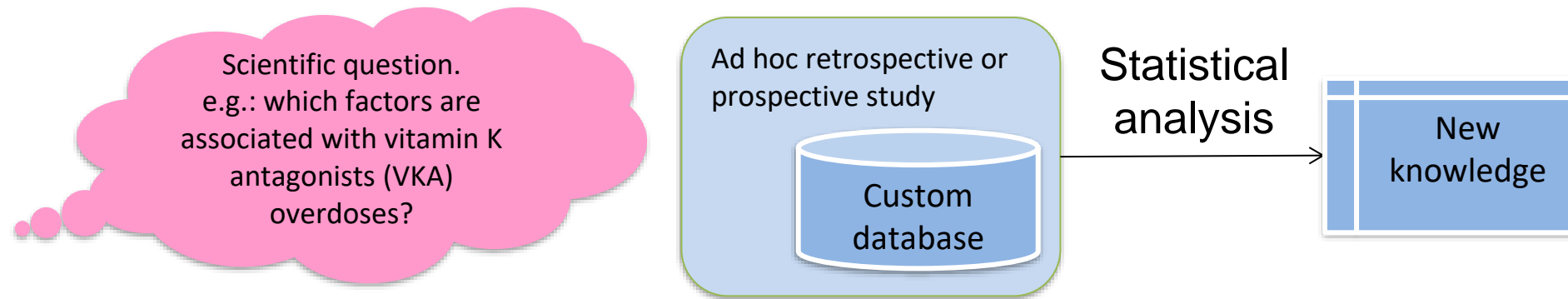




Data reuse of patient records

Definition of Data reuse (or secondary use of data)

Traditional approaches (before data reuse):



■ Advantages:

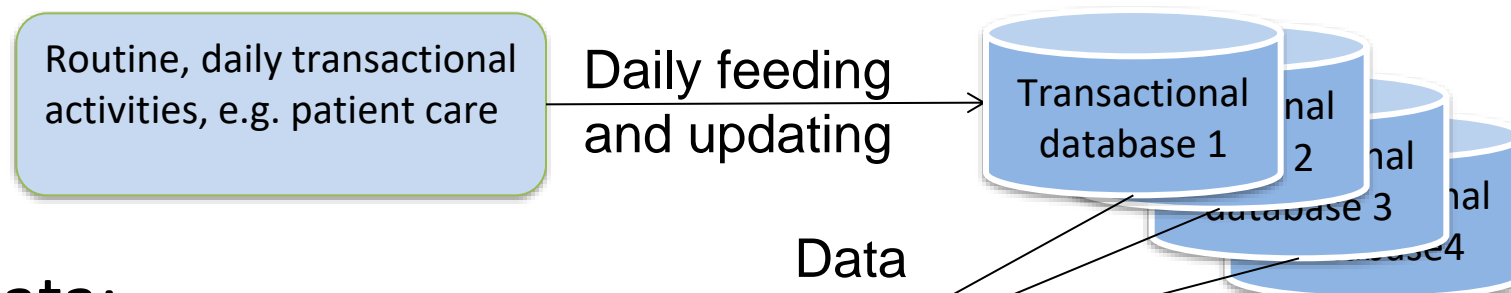
- Simple and specific data collection
- Simple data analysis
- Answers accurately the initial question

■ Drawbacks:

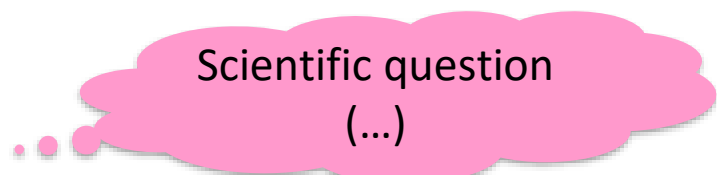
- Time-consuming
- Expensive
- Late results
- Often few records (low power)
- Data cemeteries

Definition of Data reuse (secondary use of data)

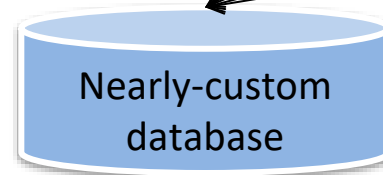
Routine collection of transactional data:



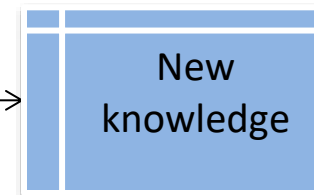
Reuse of the data:



Data transformation



Statistical analysis



■ Advantages:

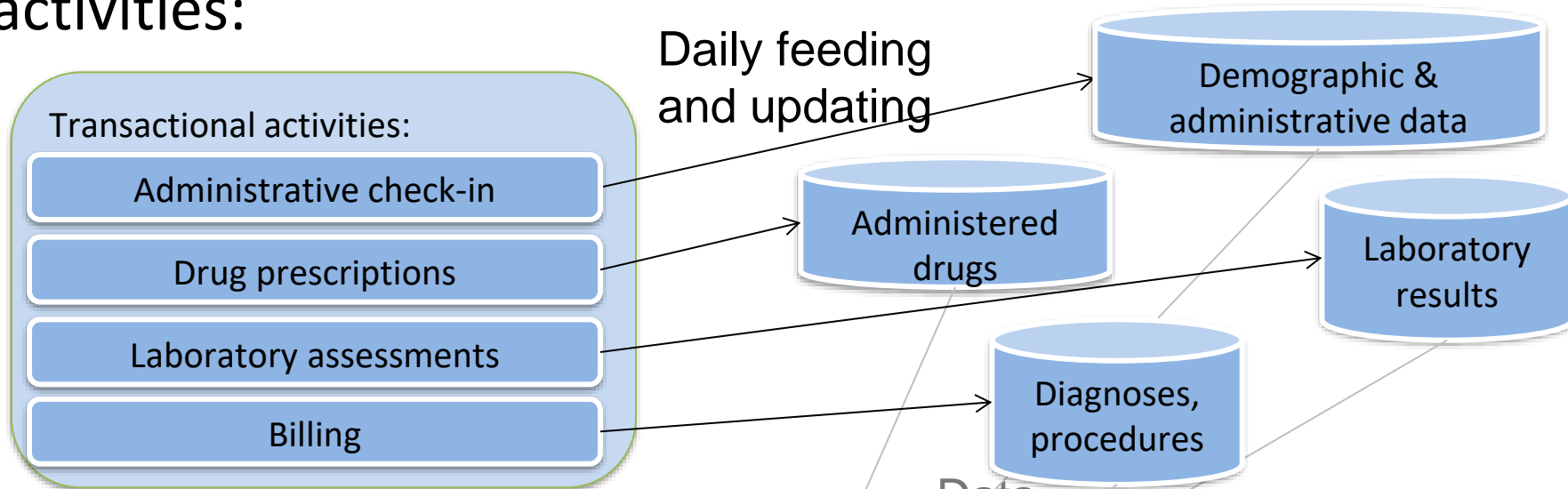
- Low-cost
- Fast results
- Data enhancement
- Amount of records => high statistical power

■ Drawbacks:

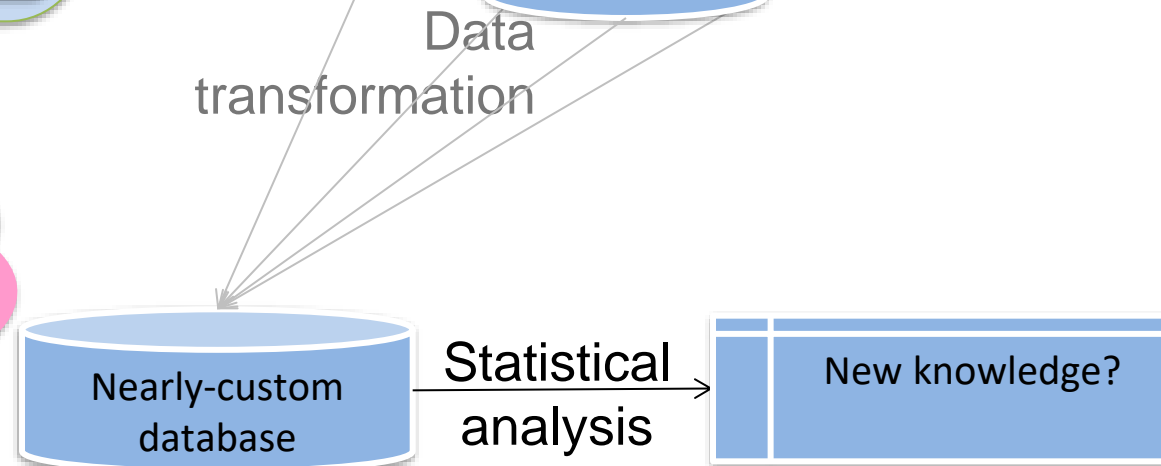
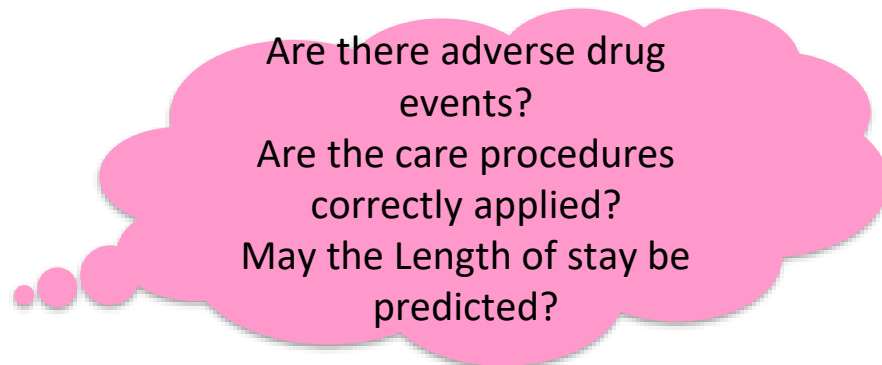
- Often approximately answers the question
- Not easy, methodological issues

Data reuse in healthcare research

Routine activities:



Reuse of the data:





Big data

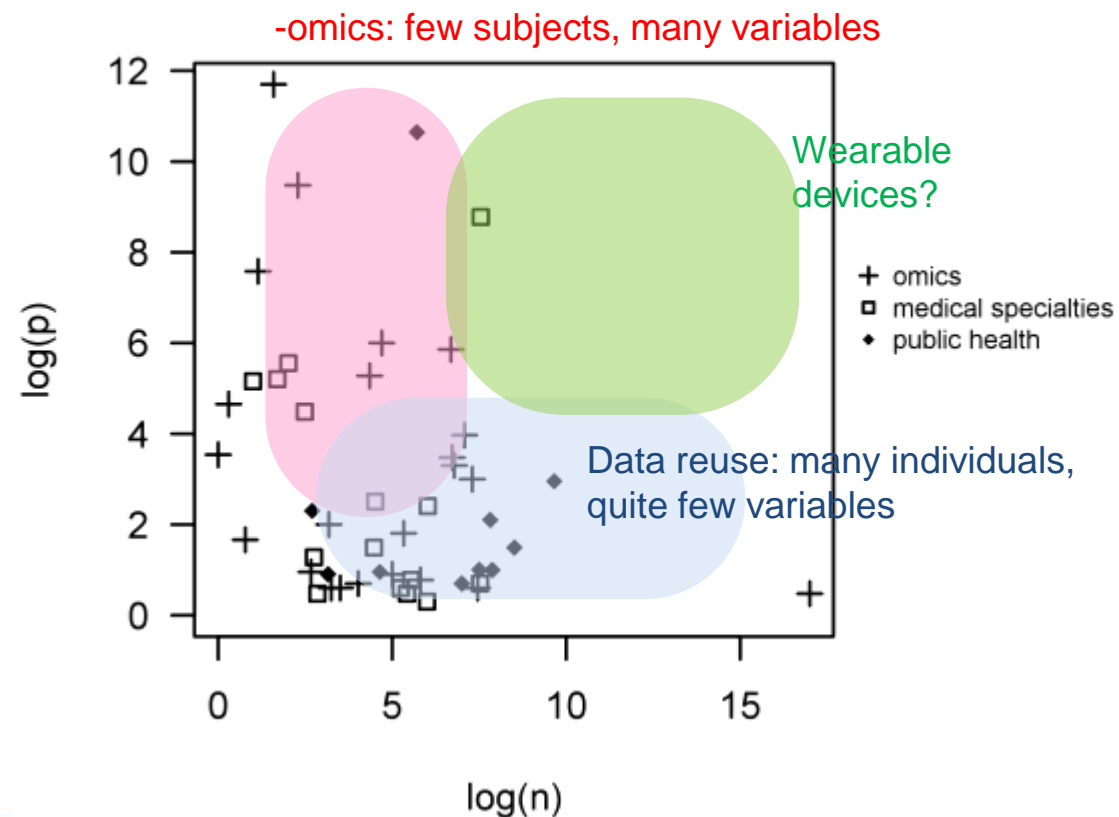
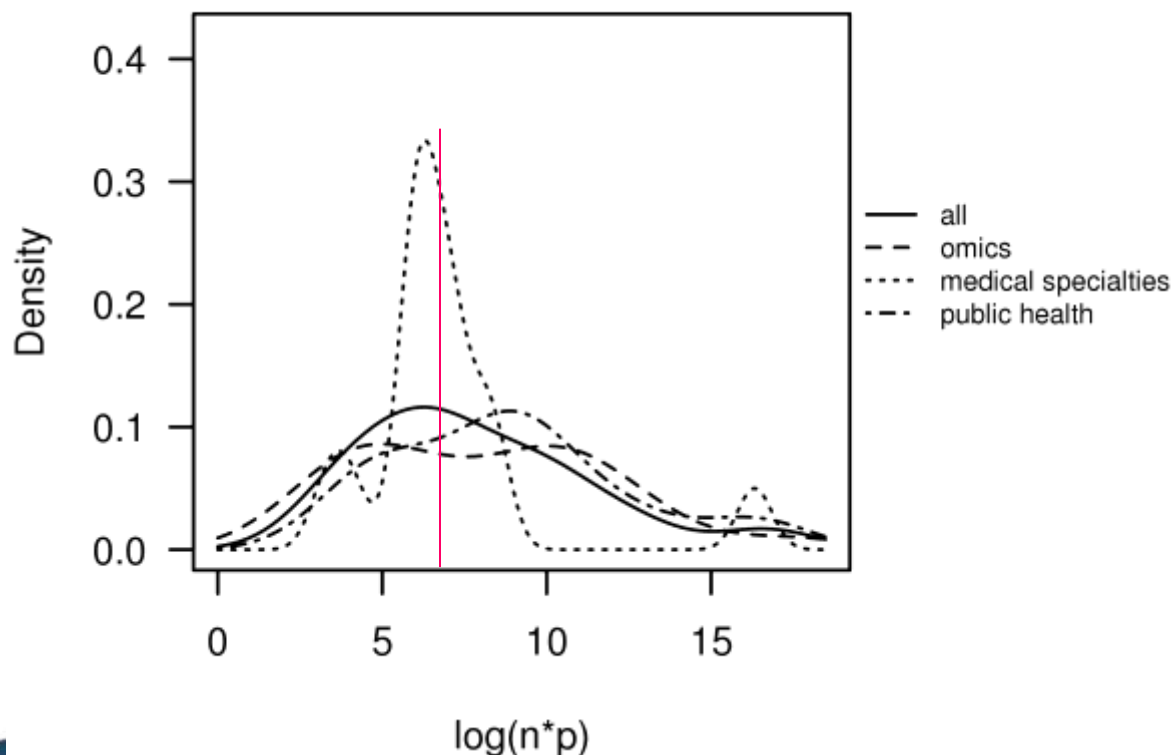
Definition of Big Data in Healthcare

Baro 2014, Toward a Literature-Driven Definition of Big Data in Healthcare

Literature-driven definition of big data :

n =nb of subjects, p =nb of variables

high volume: $\text{Log}(n \cdot p) \geq 7$



Healthcare big data

Big data sources:

- Medical records (electronic health records)
- Laboratory results (~100 lines per inpatient stay)
- Medical imaging
- Drug prescriptions (~20 lines per inpatient stay)
- Medical insurance
- Lille University hospital: 2 million patient records

Some examples in France:

- 1.5 millions diabetic patients
- 60 millions records for Social Security
- PMSI: ~30 million acute care inpatient stays per year

Definition of “big data”

“big data” is generally a property of the routinely collected data that can be reused
“Big” can be understood through 5 dimensions:

2-Many variables

1-Many records

id	age	gender	diagnosis	...
123	23	M	I10	...
125	78	M	K37	...
245	13	F	M61.2	...
278	24	M	I41	...
324	65	F	I48	...
350	34	F	F20.2	...
...

3-Many possible values for qualitative variables

4-Many tables & relationships

...
...
...
...

5-Variables with repeated measurements

Id	Par	Val
123	K+	4.5
123	K+	4.8
123	K+	5.2



Artificial intelligence

Some definitions

Intelligence:

- all the mental faculties needed to:
 - understand things and facts
 - discover relationships between them
 - arrive at conceptual and rational knowledge
- Detected through:
 - ability to understand (retrospective)
 - Ability to predict, and then to adapt easily to new situations (prospective)

Artificial intelligence:

- research into ways of endowing computer systems with intellectual capacities comparable to those of human beings, or at least mimicking them:
 - analysis (retrospective)
 - decision-making (prospective)
- but artificial grass is not grass...

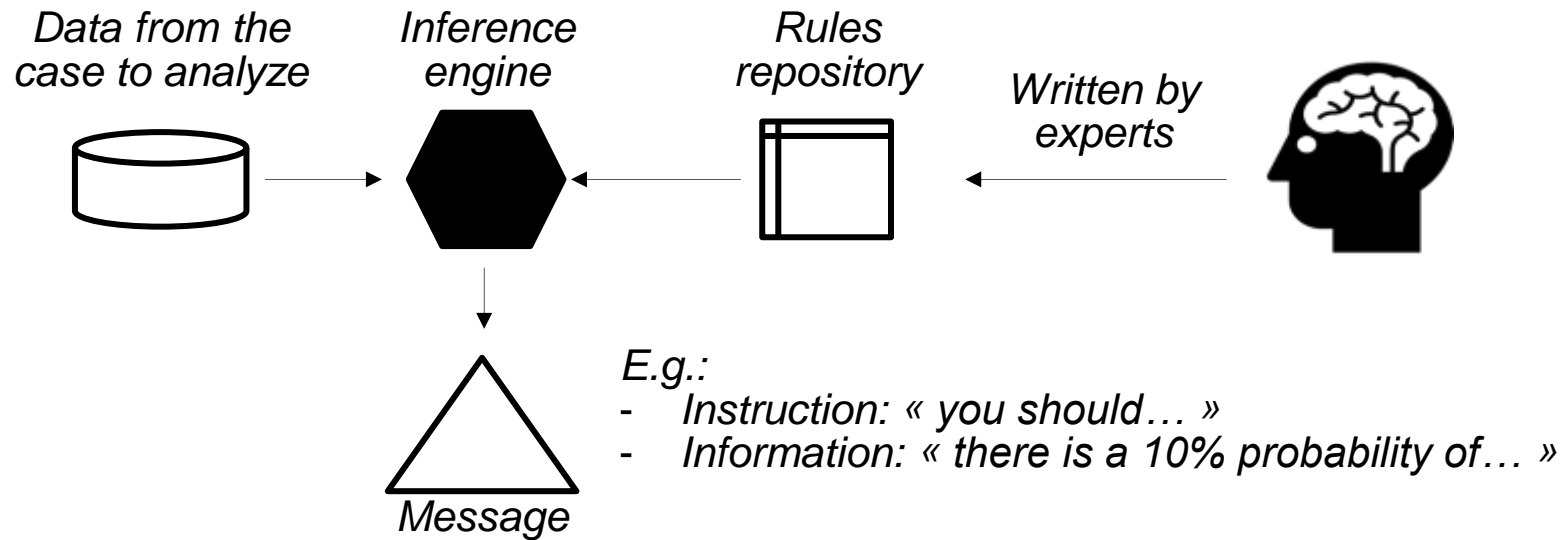
Level 1: Logical-symbolic AI = algorithmic AI

Execute a set of rules

Rules repository designed by experts

Inference engine: identifies when rules apply

=> enables rules to be applied as planned



Level 2: Data-based AI = machine learning

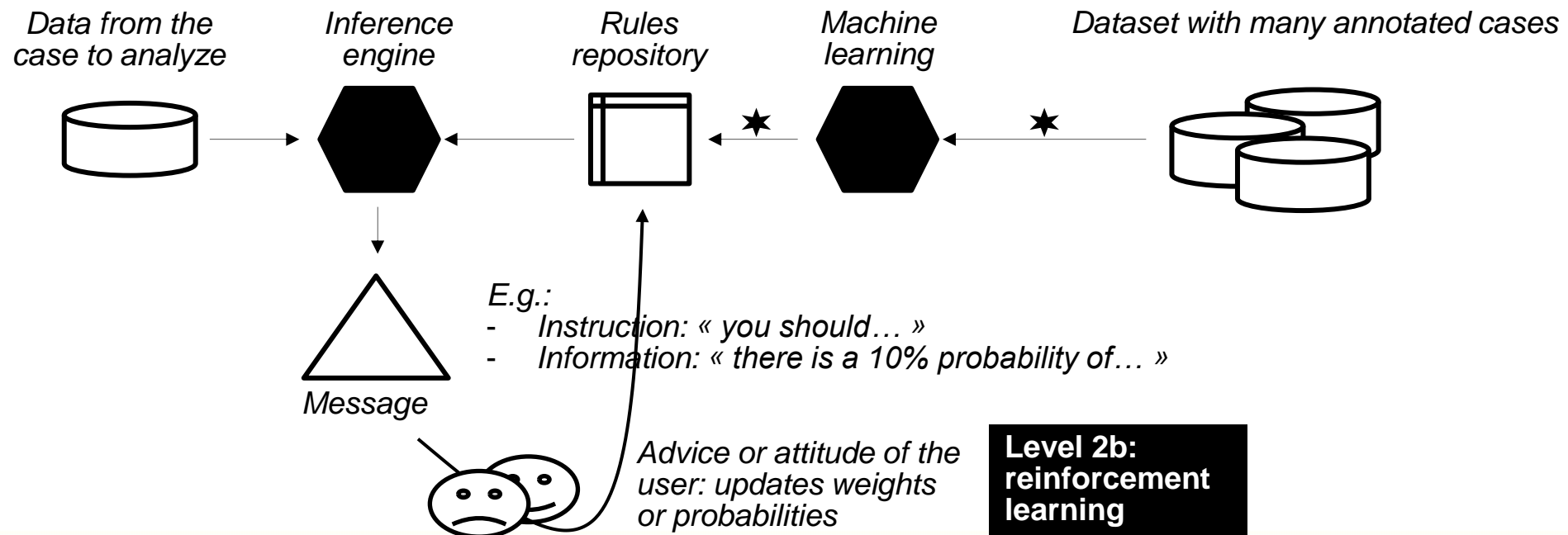
Automatically write, then execute a set of rules

Preexisting database

Statistical associations are automatically learnt (machine learning), then written as a ruleset

Then...

=> enables to manage an already known situation



Conceptual levels of artificial intelligence

Level 3: Self-learning systems

Learn from a auto-generated dataset

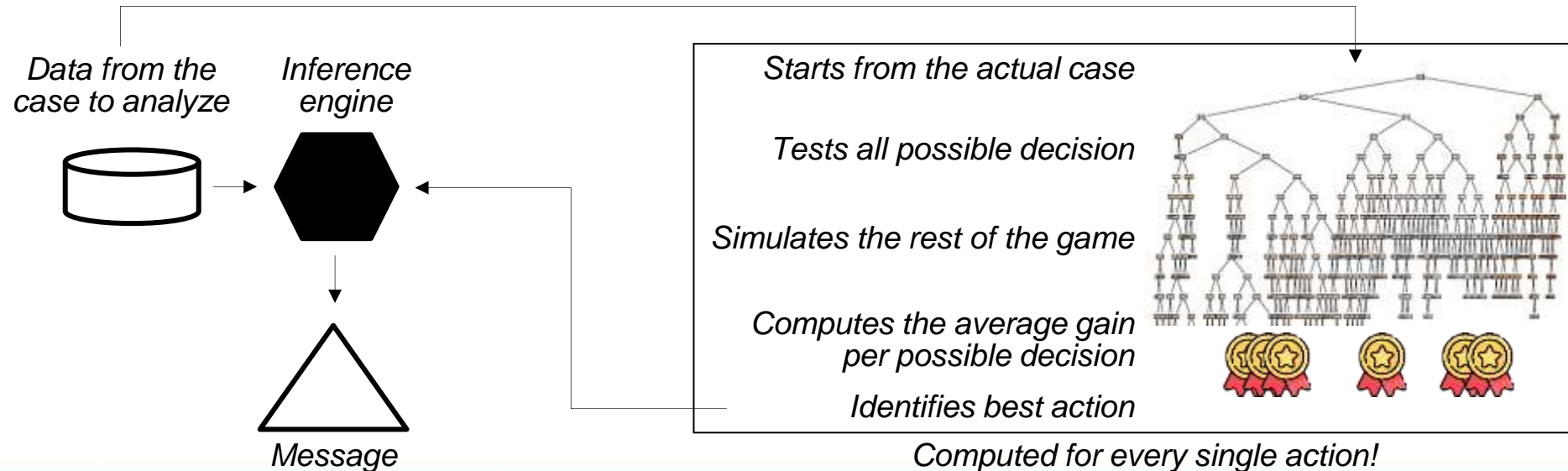
Implements the case as a starting position

Tests all the possible decisions

Simulates the next steps, and computes an average reward

Identifies the best option

=> only in a finite world, and when a gain can be computed (e.g. video games, board games, etc.)



Successful AI solutions

Clever mix of AI strategies

- E.g. automated game player
 - Logico-symbolic AI to describe the universe (game rules)
 - Self-learning system to find the best action
- E.g. classification in natural language processing:
 - Self-learning systems to pre-train the learner
 - Machine learning to train the learner
 - Logico-symbolic AI to deal with negations

Clear definition of the use case

- Use the AI module where it works best
- Evaluate the AI module to its advantage ;-)

Brief history in Healthcare

- Definition of fundamentals
 - Theoretical definition in the 1950s
 - Perceptron and artificial neural network in the 1960s (computer modelling of neural function)
- 1970s: first winter of the AI (disappointment)
- Late 1970s: Generalization of expert systems (CDSS clinical decision support systems), AI level 1
- Late 1980s: Second winter of the AI
 - Deletion of the term AI, but continued development of CDSS
- 1990s: Level 2 AI
 - concept of data mining, machine learning, knowledge discovery in databases
- 2000's: New AI boom
 - New machine learning methods: deep learning, multi-layer neural networks, support vector machine
 - No new concept, but ubiquitous AI (smartphones...)
 - 2015: Politicians and journalists discover the AI and talk about it in the future...



Explainable Artificial Intelligence (XAI)

Need for improving accountability, and critical sense of users

Methods to enable humans to understand the mechanisms of AI, and then to anticipate potential errors

Level 1: logical-symbolic IA:

- always explainable
- need for smart presentation

Level 2: machine learning:

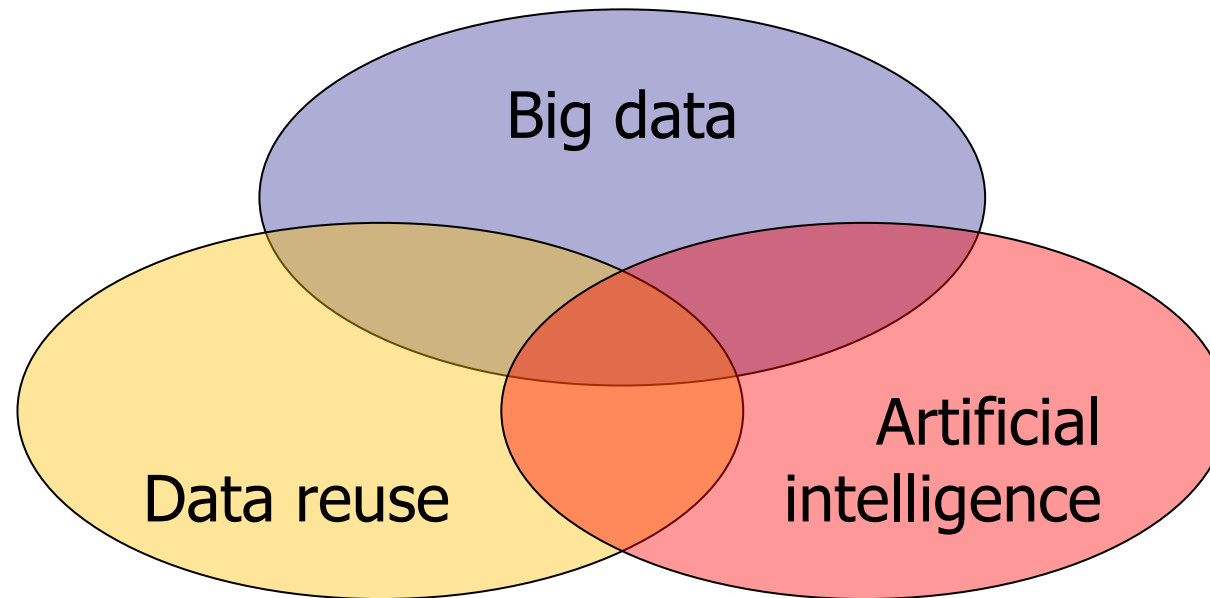
- Explainable when “white box” methods are used, e.g., logistic regression, decision trees, etc.
- Not explainable when “black box” methods are used, e.g., deep learning, SVM... but methods exist to improve explainability



Data reuse, big data, and artificial intelligence...

Data reuse, big data, and artificial intelligence...

E.g.: the French Ministry of Education disseminates the results of the baccalaureate of 743,000 candidates



Ex : a journalist team shows that:
 $P(\text{success} / \text{FirstName} = \text{'Joséphine'}) = 97\%$

Ex : an online form asks your first name
and predicts your success probability



Example of automated ECG interpreters

group	varname	kappa
Global	normal variant	0.004
	error nonfatal	0.098
	permanent pacemaker	0.195
A Morph	left ventricular hypertrophy	0.382
	right ventricular hypertrophy	0.111
A Isch	myocardial infarction	0.267
B Cond	deg1 atrio ventricular block	0.499
	left bundle branch block	0.551
	right bundle branch block	0.706
	wolff parkinson white	0.499
B Repol	repolarization abnormality	0.04
B Rhythm	sinus rhythm	0.428
	sinus bradycardia	0.441
	sinus or supravent tachycardia	0.526
	atrial fibrillation flutter	0.66
	multifocal or ectopic atrial rhythm	0.31
	prem atrial contraction or supraventricular extrasyst	0.27
	junctional rythm normal or acc	-0.002
	ventricular extrasyst	0.401
C	tachycardia	0.56
	bradycardia	0.441
	pr short	0.081
	axis deviation	0.547
	qt long	-0.002

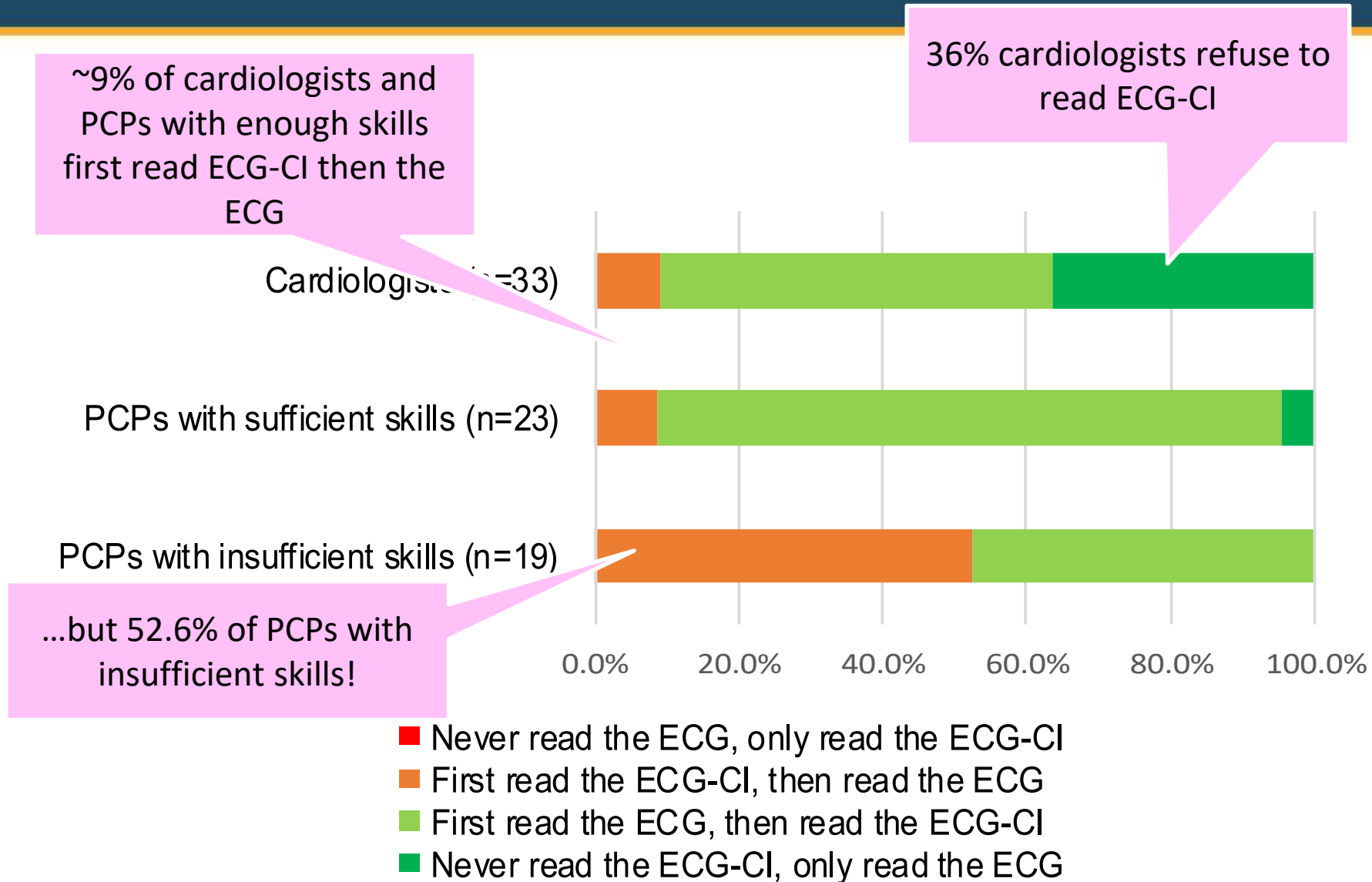
Validation study of a commercial ECG-CI

Agreement between the software and the gold standard (3 cardiologists), study of 900 records

Kappa coefficient:
 -Random answer: $k=0$
 -Full agreement: $k=1$

Is the software reliable? NO!

Usual attitude of physicians having ECG-CI



Willem Einthoven, in 1906: ECG transmission via telephone lines, 1500m,



Remote ECG interpretation



Example of



In 2014 already...

- 1000 remote healthcare centers
 - 1 general practitioner, 2-3 nurses, etc.
 - ECG recording
- Web transmission
- 1 unique center for interpretation. Every day:
 - 3 on duty cardiologists
 - about 1,000 ECGs are interpreted!
 - average delay: 3 hours



Thank you for your attention!

Emmanuel Chazard

emmanuel.chazard@univ-lille.fr

<https://www.youtube.com/@emmanuelchazard>