NOVA MEDICAL SCHOOL

CHRC COMPREHENSIVE HEALTH RESEARCH CENTRE

IPST Instituto Português do Sangue e da Transplantação, IP

APPROVED 2025 APPROVED

SUMMER MEDICAL SCHOOL: LIFESTYLE

presentation on BIG DATA & ARTIFICIAL INTELLIGENCE IN HEALTHCARE

| Guest Speakers |

Rúben Araújo – *Mechanical & Biomedical Engineer*

Luís Ramalhete – *Scientific Director of IPST Serology Laboratory*

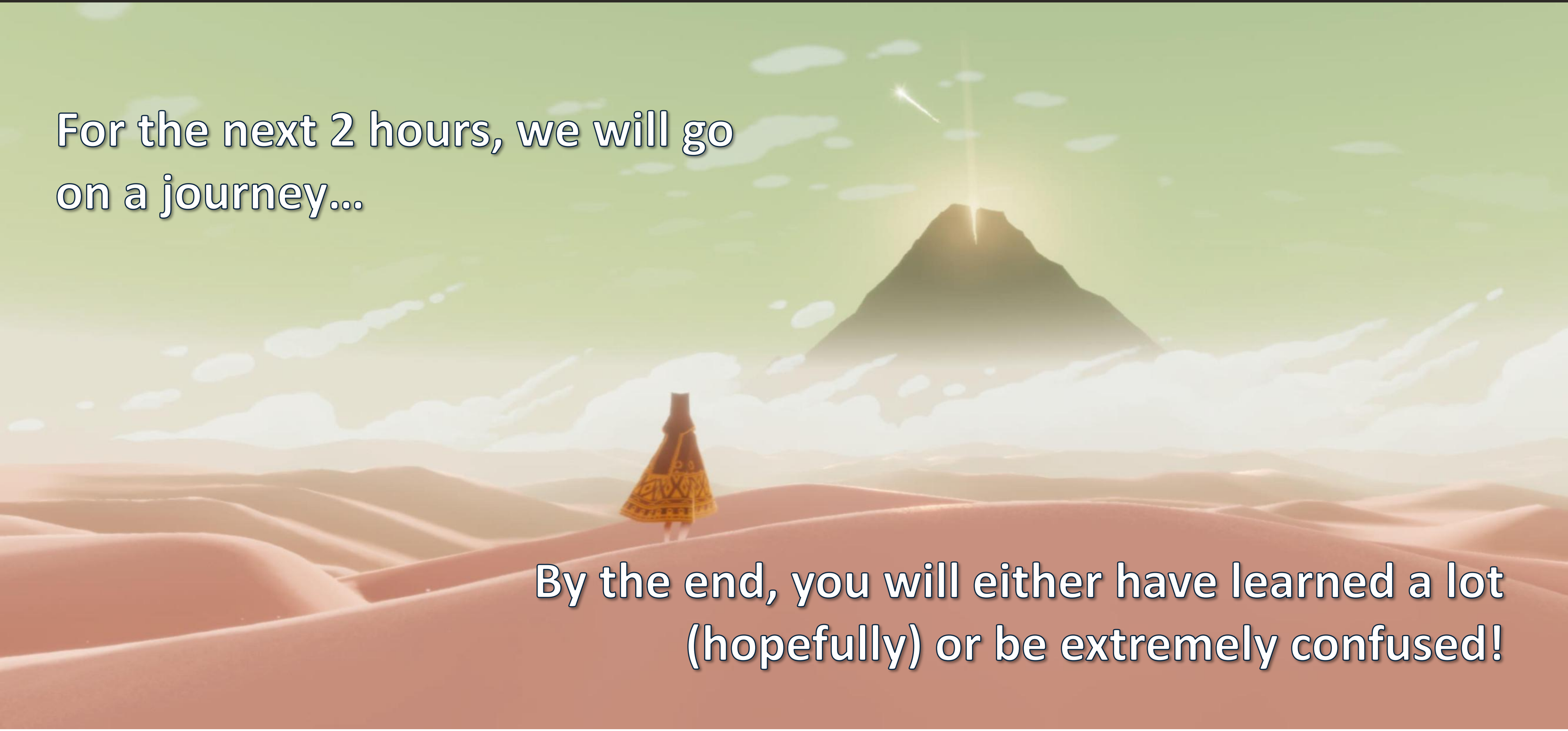
Lisbon, 3rd July 2025

Someone called me?

We were asked do discuss Big Data & Artificial Intelligence in Healthcare…

But we will go nd u a small "Crash" cours n wha needs to be a data scientist/researcher too!

So many pop culture references…so little time!



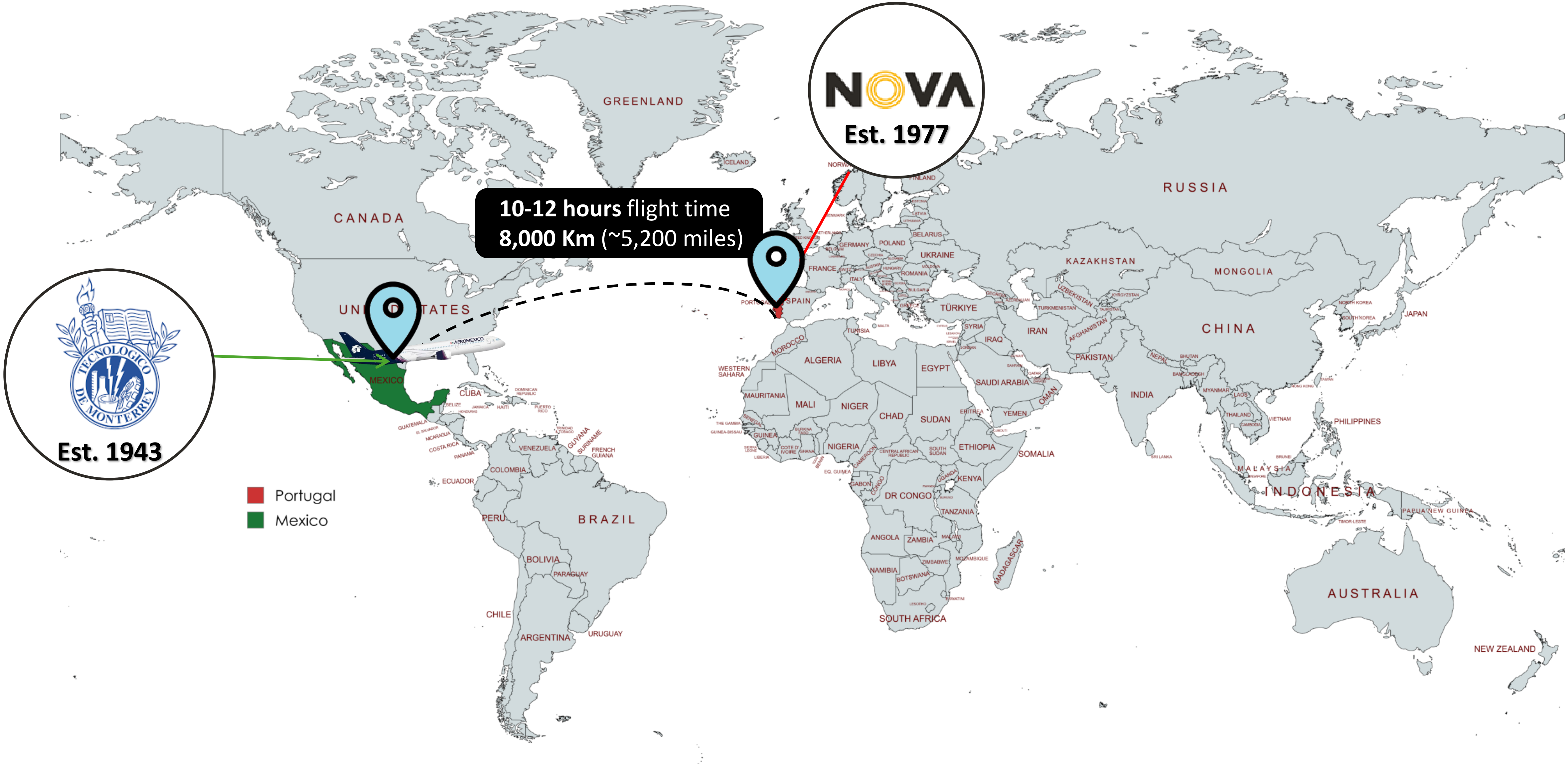
For the next 2 hours, we will go on a journey…

By the end, you will either have learned a lot (hopefully) or be extremely confused!

But first things first…

🎉 Welcome Everyone! 🎉

You sure have come a long way!



01

A Number Act

A brief summary of data between Mexico and Portugal and respective institutions and its history.

02

Who are We? And You?

A little introduction on your hosts for the next two hours, and a little discussion on what you, the students, hope for in the future.

03

A's and the…B's

Here, we will discuss a bit on how the difference fields of S.T.E.M are related and vital for any aspirant biomedical science and medicine wizard.

04

Getting Technical

The basics of each technology to be discussed is briefly presented in a easy digestible fashion.

05

Biomarkers

What is a biomarker, why is it important, and what can it help us achieve?

06

Machine Learning/AI

The tools of the trade are discussed here. What is typically used these days, what is considered the state of the art, and what does the future hold for medicine empowered by AI tools.

07

Organ Transplantation

An insight into what is done in this emerging field and what technologies are used and how does it impact clinical care.

08

Lessons Learned

What have we learned? Takeaway lessons and did your view of anything change, and if so why?

01 A Number Act

…on some curiosities…

A comparison of sorts…

Estados Unidos Mexicanos
(informal name: Mexico)

República Portuguesa
(informal name: Portugal)

- **Monterrey, Nuevo León**
- **City population:** ~ 1.14 million
- **Metro / Urban population:** ~ 5.27 million (3rd largest)
- **Tecnológico de Monterrey:** founded in 1943

- **Lisbon** (capital)
- **City population:** ~ 0.53 million
- **Metro / Urban population:** ~ 3.03 million (largest)
- **Faculdade de Ciências Médicas:** founded in 1977

Old and gold!

- Although, originally founded as the *Real y Pontificia Universidad de México* in **1551**…
- …the modern *Universidad Nacional de México* was founded in **1910** by Justo Sierra…
- …and in **1929**, the Mexican government passed a law granting the university "autonomy", gaining the new and modern name of *Universidad Nacional Autónoma de México* (**UNAM**)
- …**one of the largest** and **most respected** universities in the world.

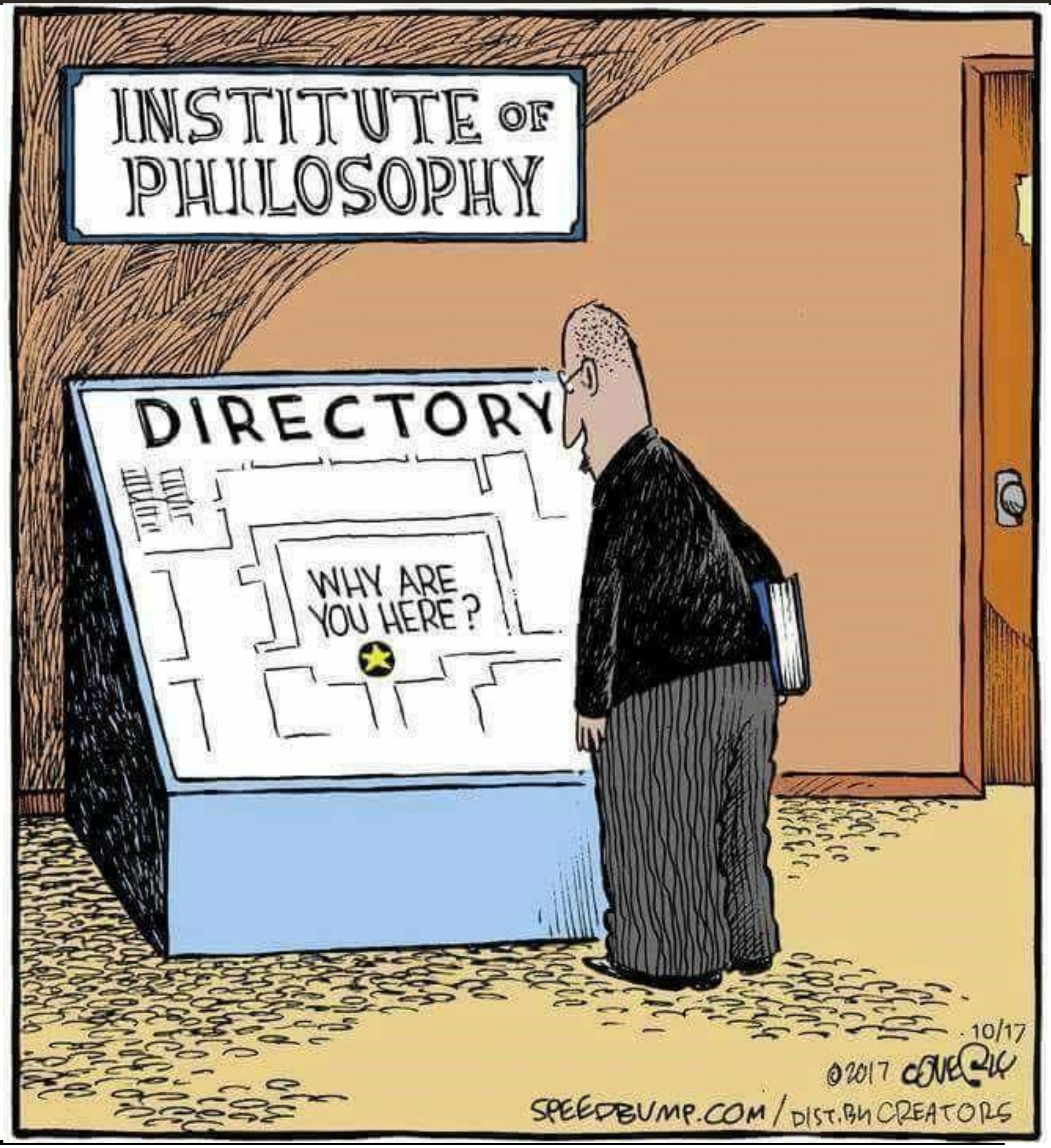
“Harry - yer a wizard!”

- **Universidade de Coimbra**, was founded in 1290.
- **One of the oldest universities in Europe** still operating (+25K students)
- It began in Lisbon, then **moved to Coimbra permanently in 1537.**
- You might recognize some inspiration behind a **famous book and movie franchise**!

Oldest Engineering School in Portugal (my alma mater)

- **Founded in 1852**, as the ***Instituto Industrial de Lisboa*** (**IIL**) during the reign of D. Maria II…
- **It was renamed in December of 1974**, after the *Revolução dos Cravos* in the 25th of April of the same year, as the ***Instituto Superior de Engenharia de Lisboa*** (**ISEL**)…
- **It’s the oldest engineering school (Polytechnical) of Portugal.**

02 **Who are We? And You?**



Who are your speakers for the next (now less than) 2 hours?

NOVA MEDICAL SCHOOL

Luís Ramalhete

Bachelor in Clinical Analyses & Public Health
Masters in Biomedical Engineering
PhD student in Biomedicine (finalist)

- Head of Technical and Scientific Operations of Serology Laboratory of *Instituto Português do Sangue e da Transplantação.*
- Specialist in machine learning and applications applied to the research of organ rejection and transplantation.

Rúben Araújo

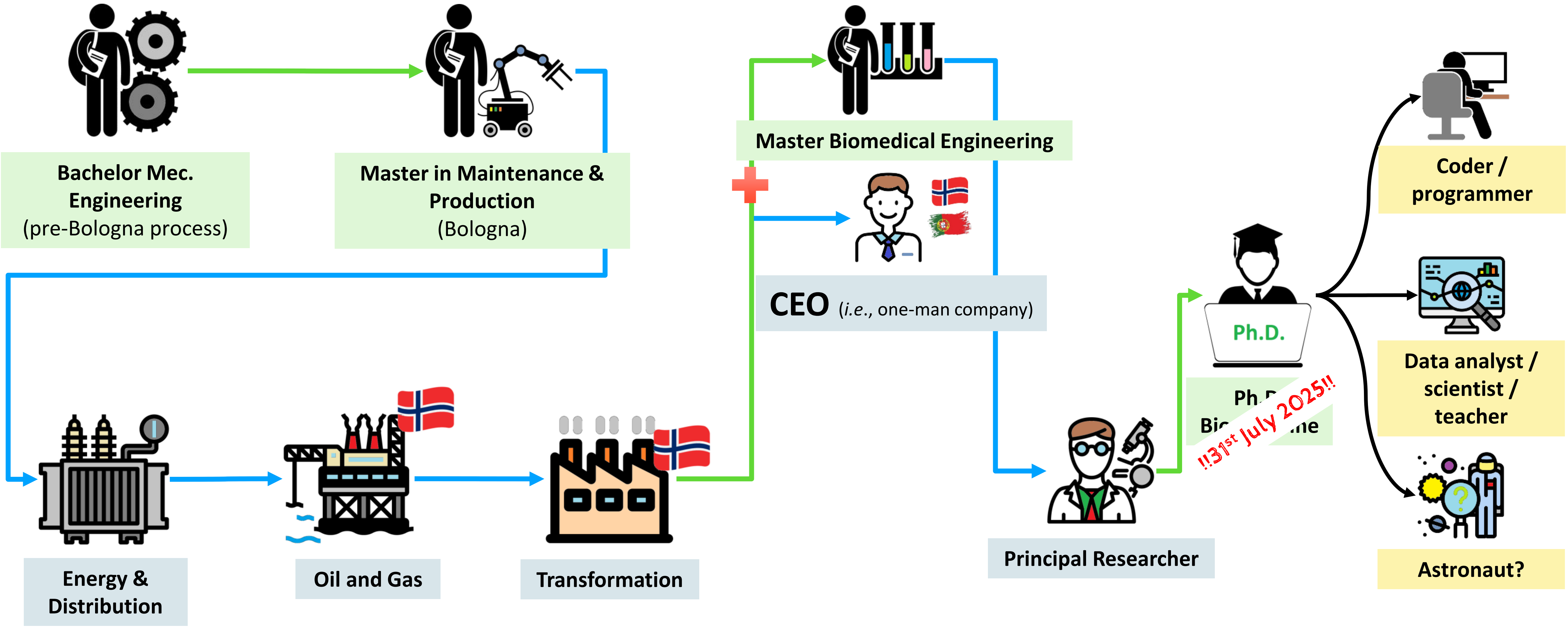
Bachelor in Mechanical Engineering
Masters in Maintenance & Production
Masters in Biomedical Engineering
PhD student in Biomedicine (finalist)

- Engineer, programmer and full-time researcher.
- Expert in metabolomics, clinical analysis, advanced statistics.

How about you? Tell us a bit about yourselves!

It’s OK to feel lost or overwhelmed still.

The path to researcher – a non-linear model

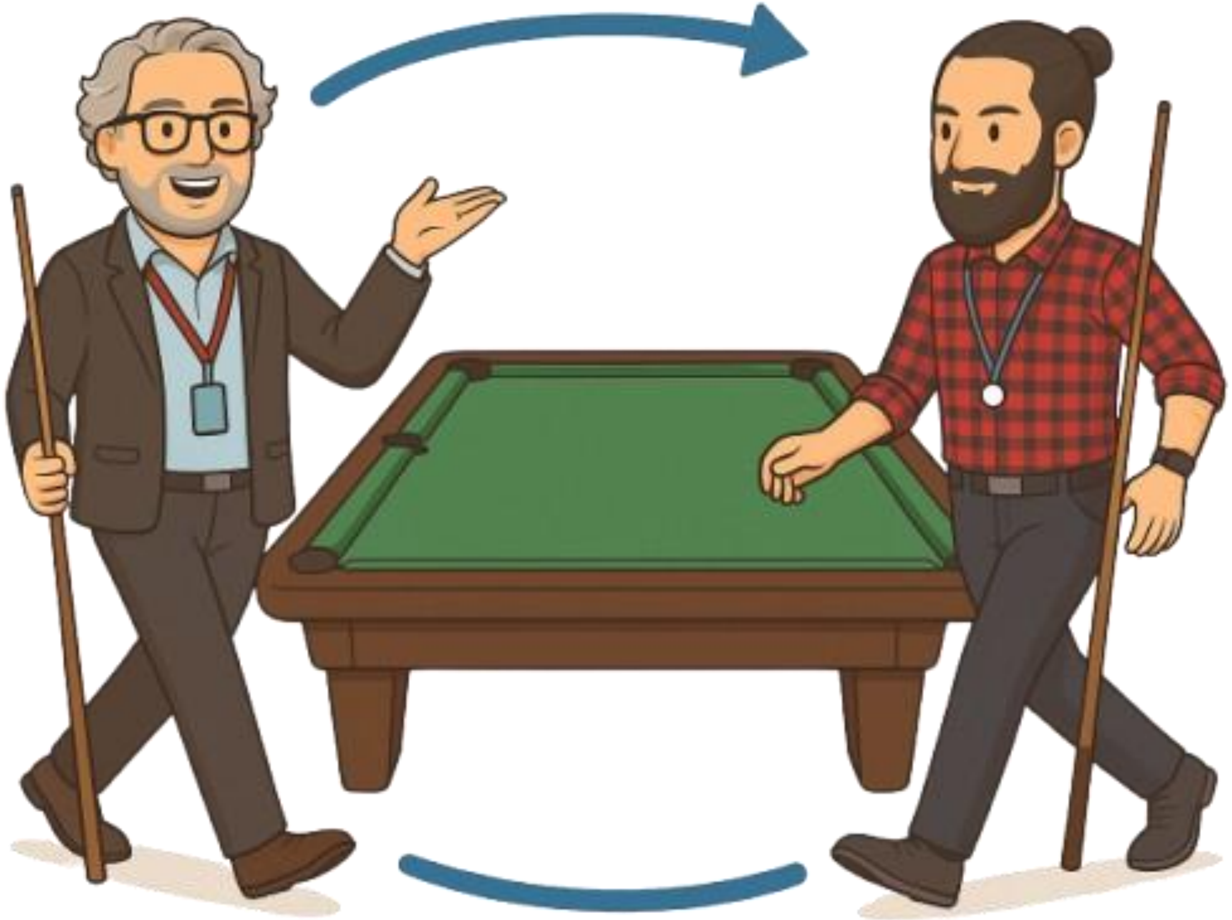


The Science behind the Madness

03 A’s and the… B’s

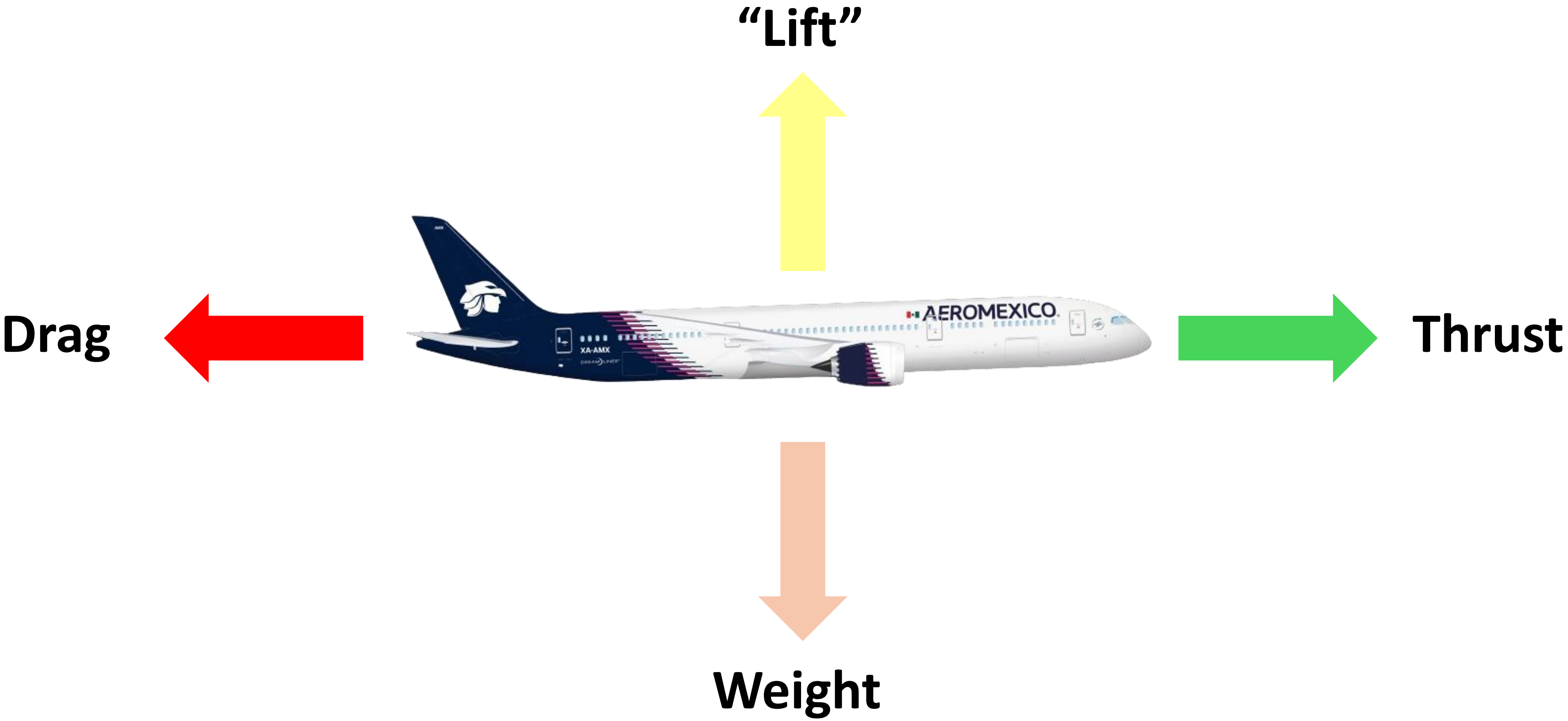
"Dar a volta ao bilhar grande"

Original meaning: an informal expression when we want someone to stop harassing us by sending them to some imaginary place.

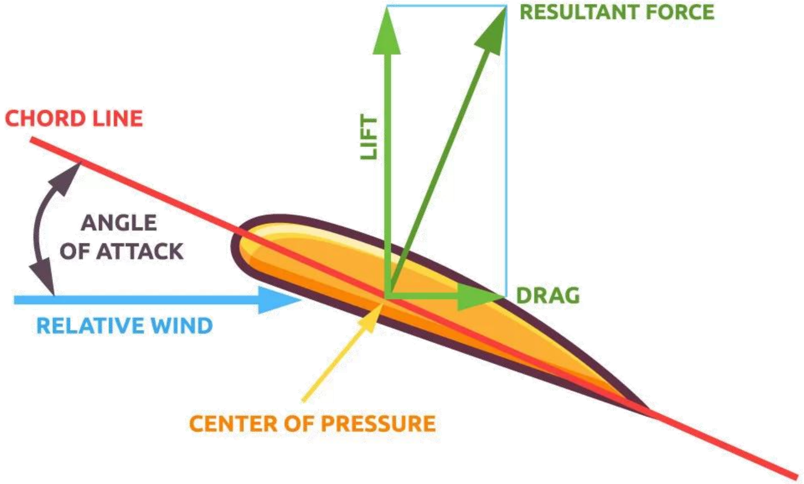


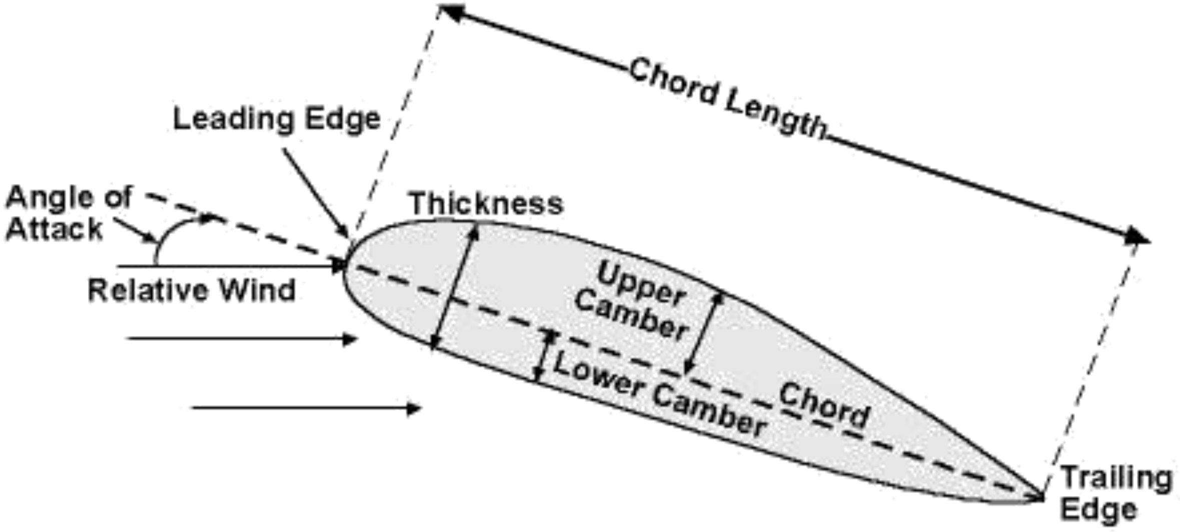
However!

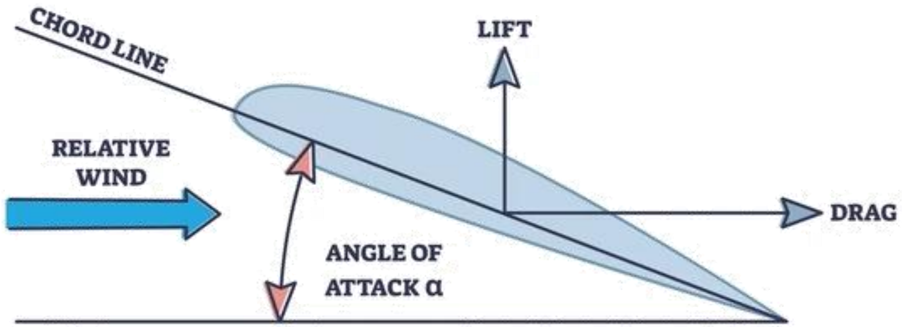
Today… we'll give it our own special twist — you'll see!


“Lift”
Drag
AEROMEXICO
XA-AMX
Thrust
Weight

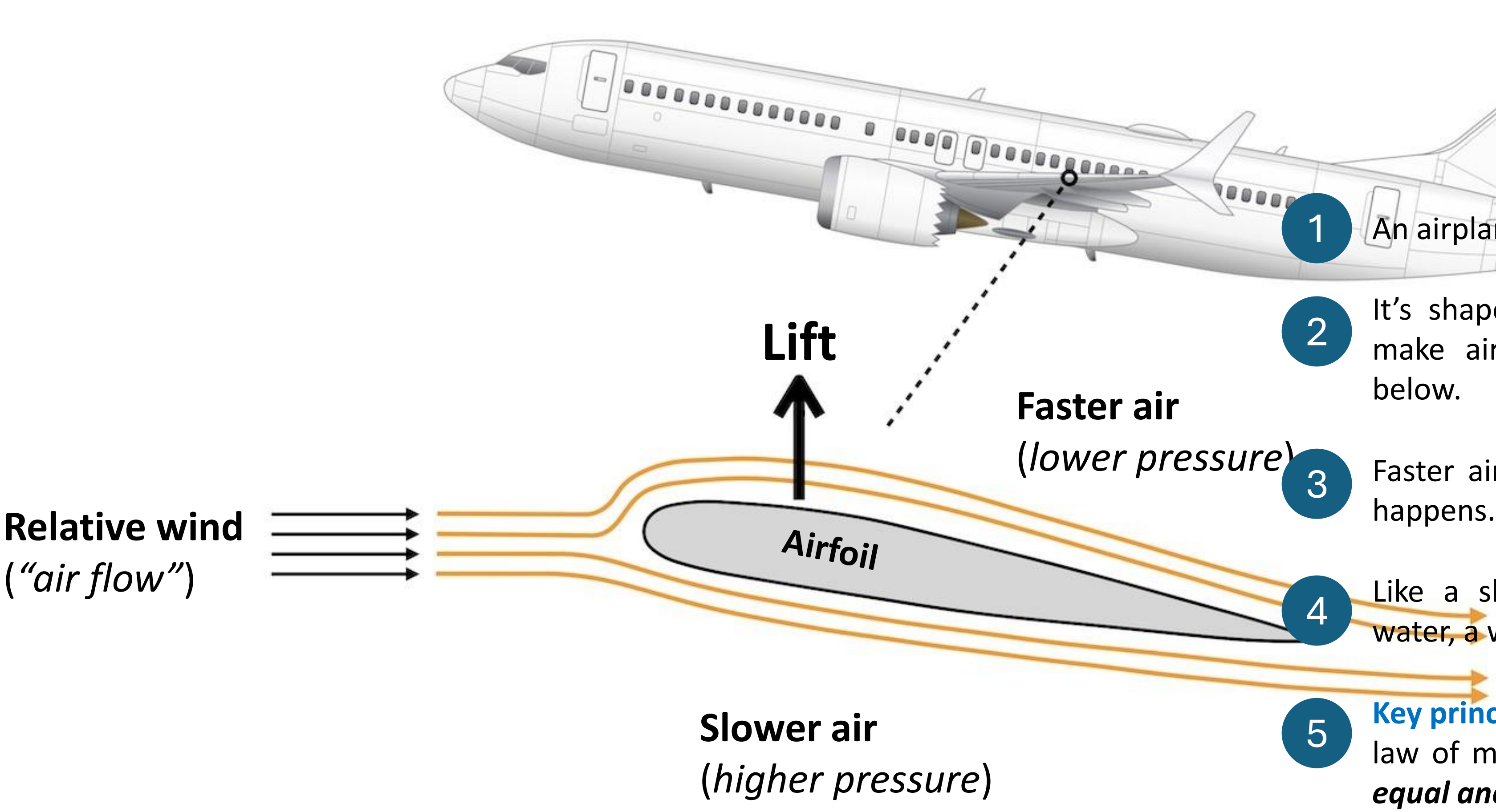
1. "Lift": Staying up in the air





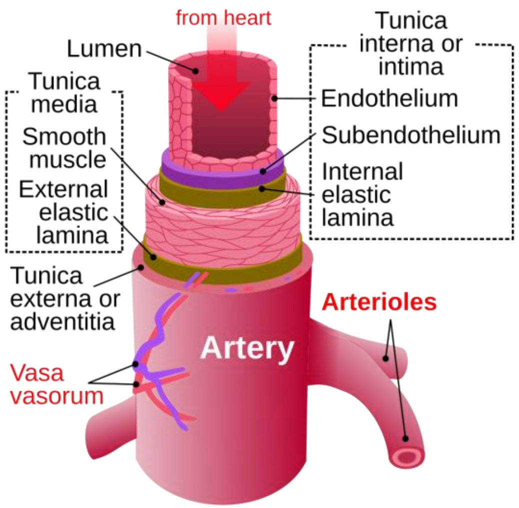


1. “Lift”: Staying up in the air



1. An airplane wing lifts by “pushing” air down.
2. It’s shape (airfoil) and tilt (angle of attack) make air move faster on top and slower below.
3. Faster air ➔ lower pressure above ➔ “lift” happens.
4. Like a ship/boat stays afloat by displacing water, a wing “floats” by pushing air down.
5. **Key principle:** Bernoulli's effect + Newton’s 3rd law of motion (“***for every action there is an equal and opposite reactio***n”).

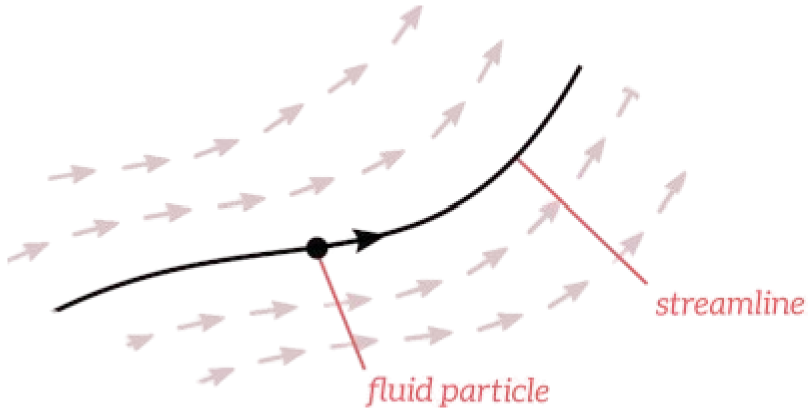
2. Blood Flow: Fluid dynamics (inside you)



Bernoulli's Principle (Daniel Bernoulli, 1738)

$$\underbrace{\overbrace{P}^{\text{STATIC PRESSURE}}}_{\text{PRESSURE ENERGY}} + \underbrace{\overbrace{\frac{\rho}{2}V^2}^{\text{DYNAMIC PRESSURE}}}_{\text{KINETIC ENERGY}} + \underbrace{\overbrace{\rho g h}^{\text{HYDROSTATIC PRESSURE}}}_{\text{POTENTIAL ENERGY}} = \text{constant (along a streamline)}$$

fluid pressure; density; elevation; velocity; gravitational acceleration

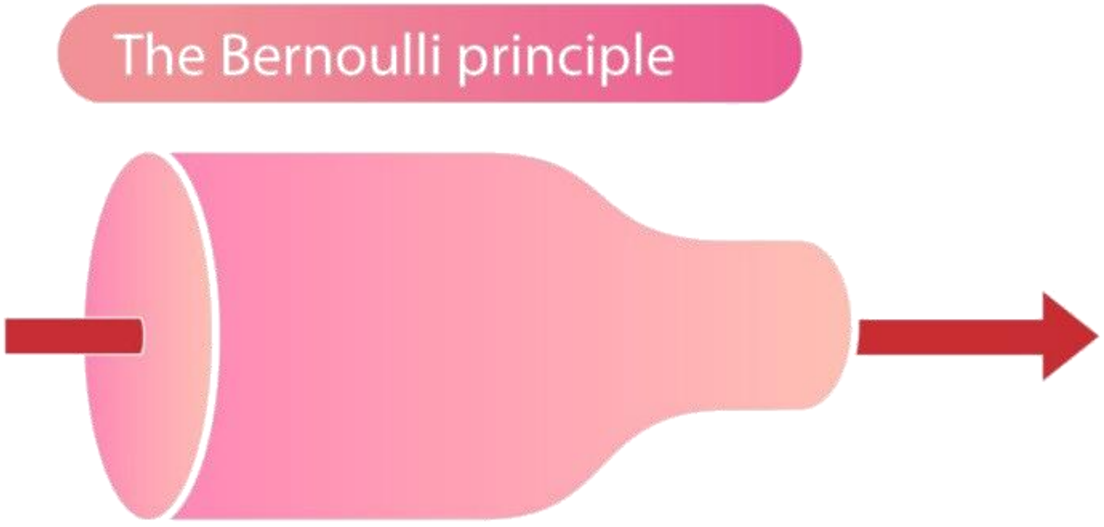


1. It describes the relationship between the **pressure**, **velocity** and **elevation** of a flowing fluid.

2. This principle states that for a non-viscous fluid in steady flow, **an increase in the speed of the fluid occurs simultaneously with a decrease in pressure** or a decrease in the fluid's potential energy.

3. *A **streamline** is defined as the path traced by a single particle within the fluid.*

2. Blood Flow: Fluid Dynamics (inside you)



1. Bernoulli's principle helps explain blood flow in arteries, particularly how pressure changes with varying blood velocity.

2. **When an artery narrows** (e.g., due to plaque buildup), **the blood flow velocity increases in that constricted area, causing a drop in pressure**.

3. <u>This can lead to the artery walls collapsing inward</u>, potentially disrupting blood flow and, in severe cases, contributing to heart attacks.

4. The vessel walls (in arteries, veins, capillaries) handle the pushing force, with tension keeping the vessel walls intact.

5. **Key principle:** Doctors use this information to understand blood pressure, blockages, and risks.

3. *Diving deep: "Under pressure" (Queen, 1981)*

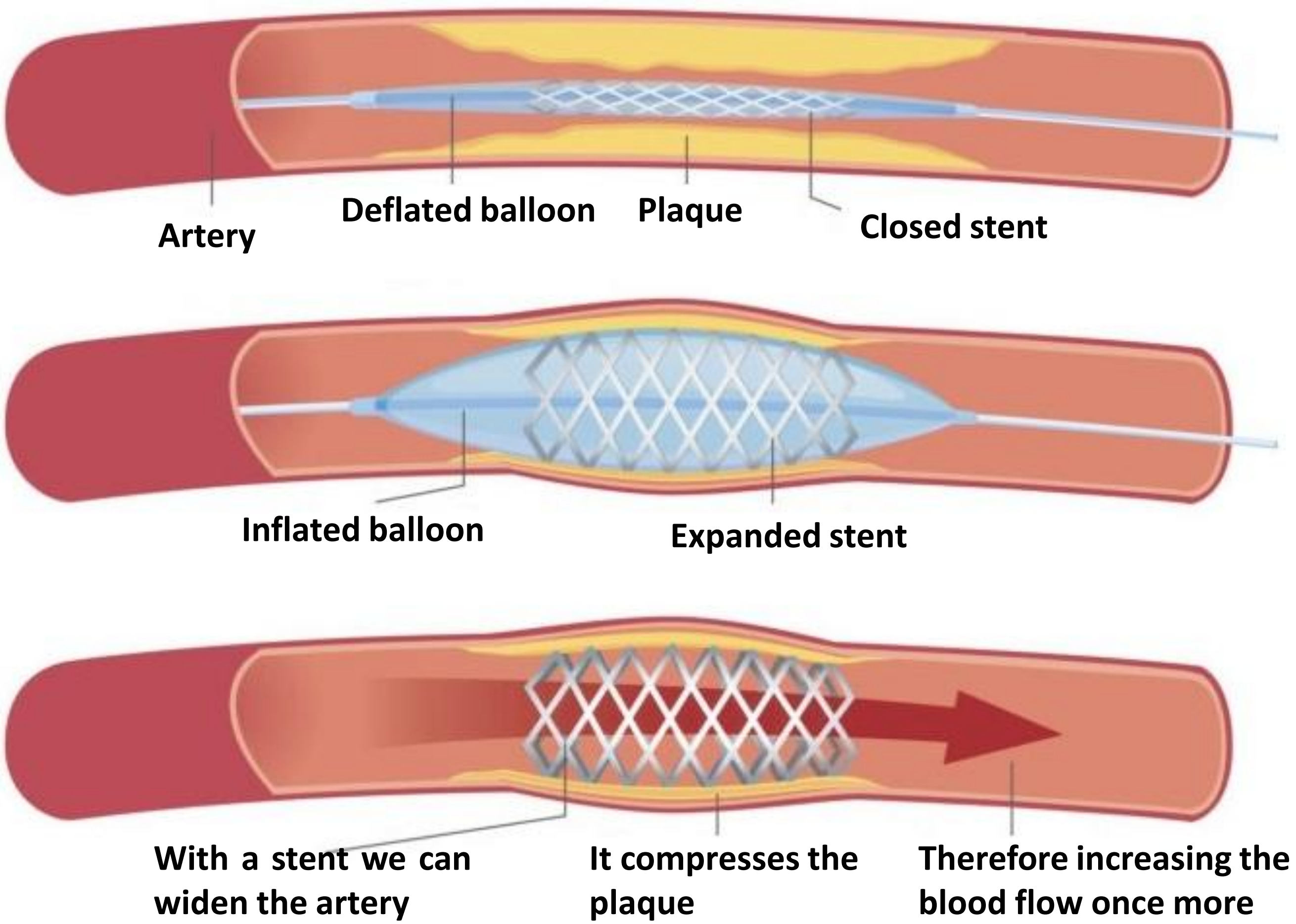
1. The standard atmosphere (symbol: atm) is a unit of pressure defined as 101325 Pa. The average pressure at mean sea level (MSL).

2. Blood shifts to the chest, vessels compress slightly, and the body adapts to keep you safe

3. "Diving underwater shows how pressure affects your body. At 10 meters deep, the water pressure doubles compared to the surface — so your lung volume halves (Boyle's Law: *inverse relationship of P and V of a gas when the temperature and amount of gas is constant*).

4. At PADI Advanced Open Water (40m) the % of lungs volume is only 20% of the original.

5. At the World Record Deep SCUBA (332 m), lung size is a measly 3% of surface size (~ size of a lime or around 90 ml volume out of an original 5-6 litres!).



1. Medical implants **use strong but flexible materials**

2. Artery stents are **typically made of metal mesh**, with common materials including stainless steel, nitinol (a nickel-titanium alloy), and cobalt-chromium alloys. Some stents are also covered with polymers, like silicone, to enhance biocompatibility or deliver medication.

3. **A stent expands inside the artery**, keeping it open and improving patient condition.

4. Not unlike the wings or other parts of an airplane, the stent must too resist **stress**, **strain**, **pressure**, and **bending**.

GLOSSARY

Stress: Internal force per unit area in a material.

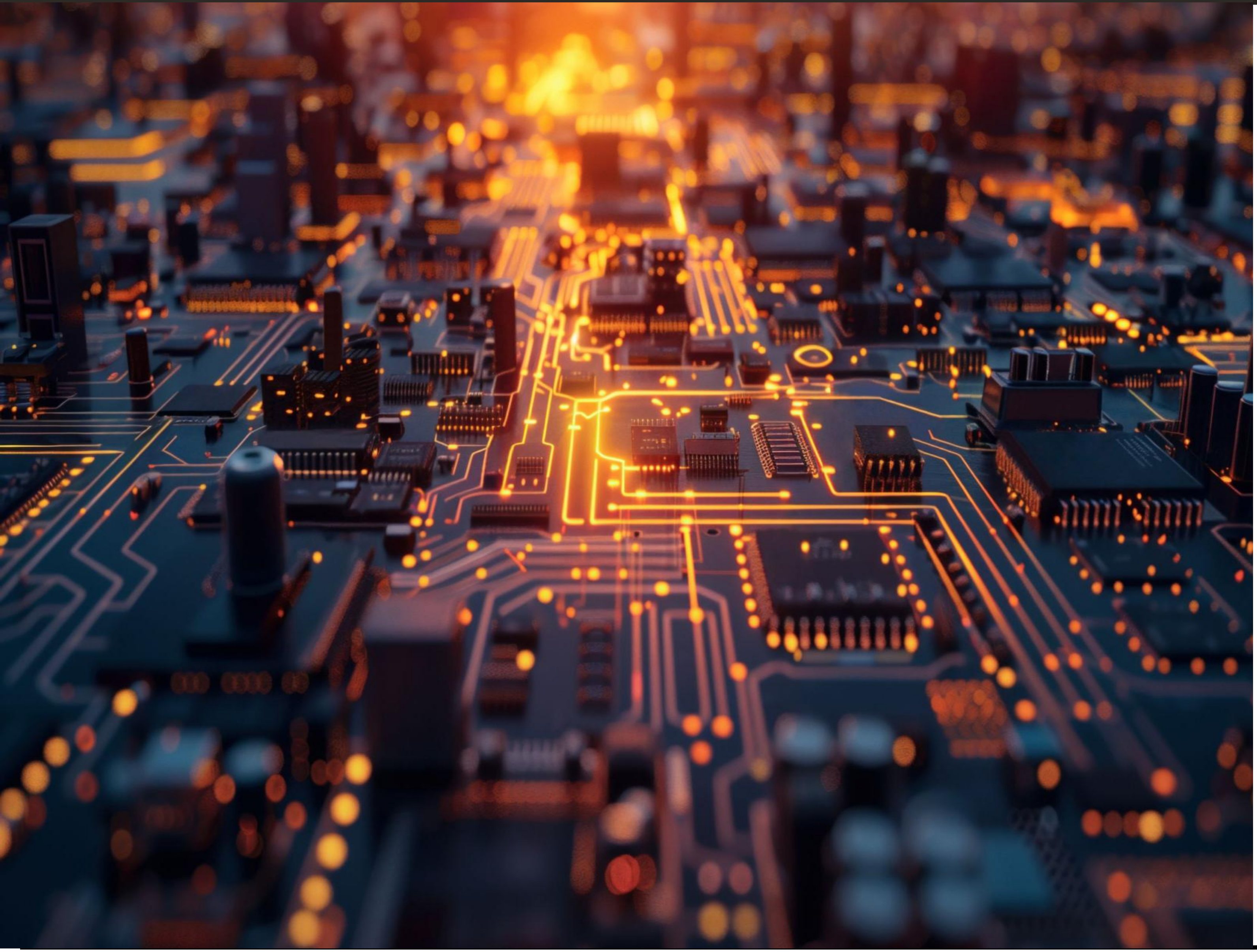
Strain: How much a material deforms under stress.

Tension: Pulling force that stretches a material.

Compression: squeezing force that shortens a material.

Pressure: force per unit inside a fluid.

5. Electronics: the brain of modern medicine



1. Devices like **pacemakers** and **insulin pumps** rely on small but complex electronic systems.

2. Complex devices can sense the body's signals and relay the data to both doctors and patients.

3. Just like aircraft sensors, here too, **sensors are used to monitor the health of a "system"** (here, the human body), 24 hours a day, 7 days a week.

6. Programming: making sense of the data

1. Implants, standard hospital equipment and even modern wearables (e.g., smartwatches, bands), **collect and process huge amounts of health related data**.

2. Through programming, “*Artificial Intelligence*” can help us **find patterns and detect the early onset of medical conditions** that can save or improve patient’s lives.

3. **Doctors, engineers, researchers** (with the collaboration of patients), all **working together** to turn raw data into better healthcare for everyone.

7. The Researcher Role: bridging all fields

1. Researchers connect **engineering, biology, electronics, and programming**.

2. They **build on past discoveries** and **push science forward**.

3. **Today's biomedical researchers need to be comfortable "*wearing many hats*".**

8. Protecting Innovation: patents

1. A patent is a "***legal document giving the holder exclusive intellectual property rights over a specific invention.***" An important element in encouraging innovation, it grants its holders exclusive intellectual property rights over a specific invention and over a specific time.

2. A patent team **includes engineers, lawyers, and medical experts** — everyone helps explain and defend the invention.

3. **Patents must be planned wisely:** companies choose where to protect them (regional or international), how long protection lasts, and budget for costs, translations, and legal steps.

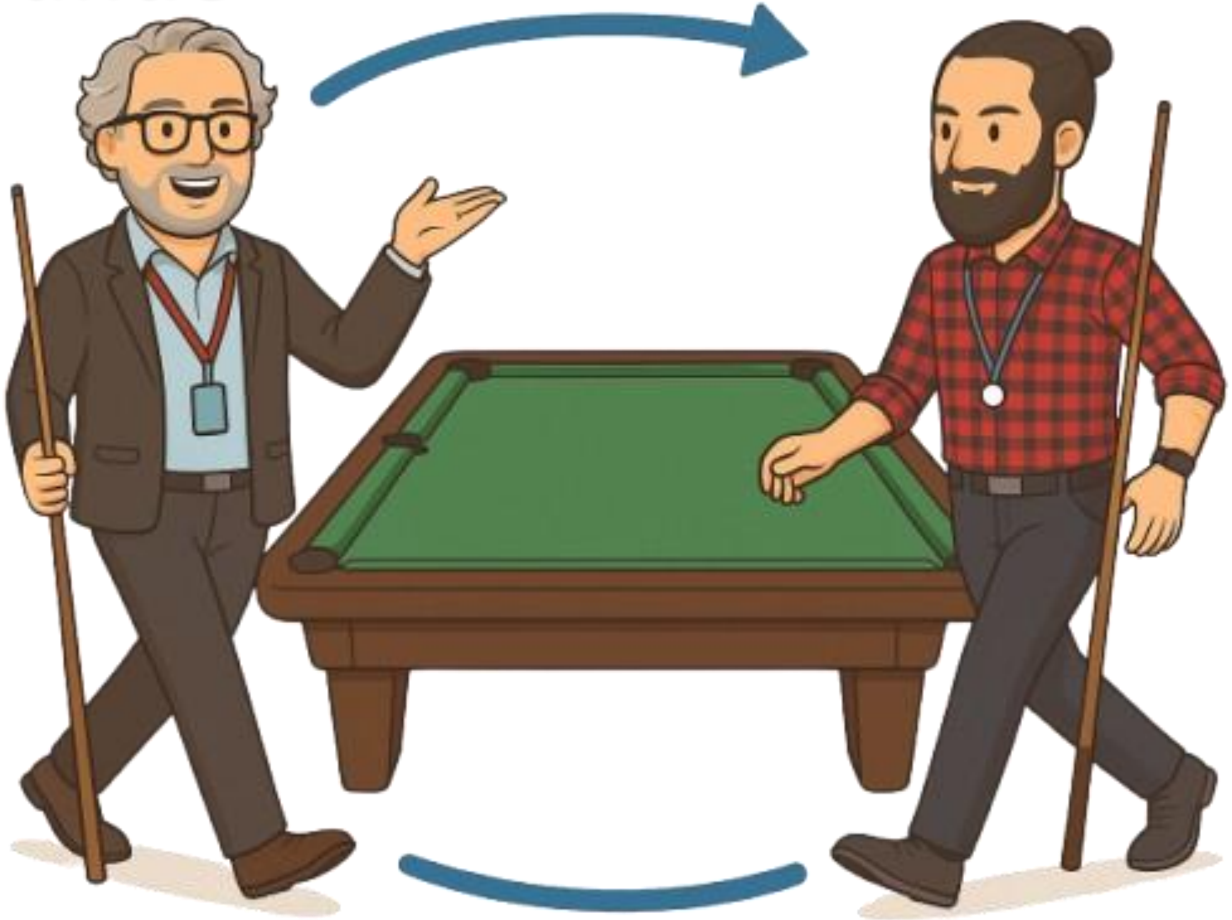
NOVA MEDICAL SCHOOL

1 **Funding turns ideas into real products** (e.g., software, hardware) **to use in healthcare**.

2 **Investors can be private** (angel investors), **public** (government), **or industry** (big pharma).

3 **Good funding + teamwork** means **faster progress and better health for everyone**!

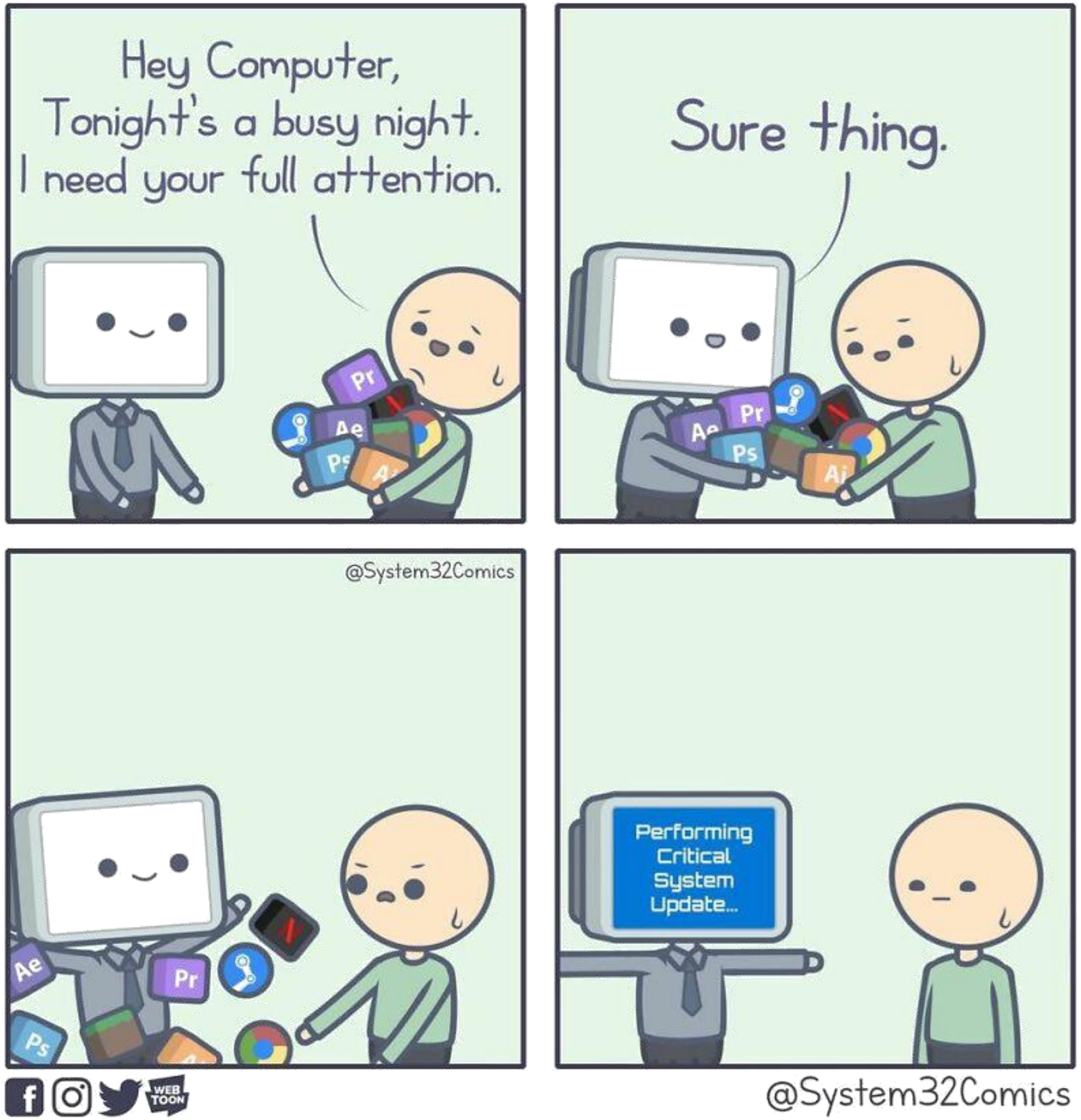
"Dar a volta ao bilhar grande"



More modern meaning: taking a long winded way to explain something, through various separate concepts that, apparently and at first, seems to have little to do with each other.

We hope you know understand there is a *long* way before getting to be a good data scientist, no matter the field you might work in one day (health, finance, space industry, etc.). **Continuous improvement and learning** is just another necessary part of being a good researcher.

So don't be afraid about your immediate career path**, as there are many ways to get to where you want to be** and that too can change over time!



04 Getting Technical

NOVA MEDICAL SCHOOL

1 Biological Assay

Basic steps of a typical biomedical assay: picking the study population, collecting biosamples, processing, and analyzing them.

2 Omics

What omics are, main branches from classic to emerging, with extra focus on metabolomics and proteomics.

3 Vibrational Spectroscopy

Quick look at what vibrational spectroscopy is, how it works, and how it supports biomedical research.

4 Cytokines

What cytokines do, difference between proinflammatory and anti-inflammatory types, and why they matter.

5 Analytical Strategies

Overview of univariate vs. multivariate analysis, and how multivariate splits into classification and regression.

Univariate analysis

Multivariate analysis: *classification & regression*

6 Classification Methods

Key unsupervised and supervised classification methods. When to pick one over the other.

Unsupervised Techniques

Supervised Techniques

7 Regression Techniques

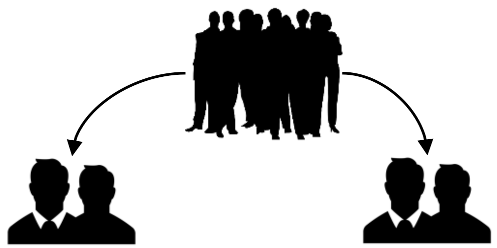
What regression methods are for, with examples of linear and non-linear types.

“Assay, therefore I am!”

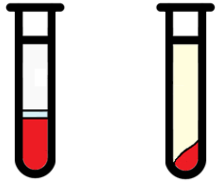
"How is a normal research assay usually conducted?"

Population → Biological Assay → Analytical Technique → Pre-processing & Processing → MVA + ML

N = 500



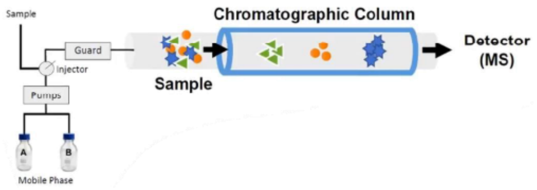
Pathology **Controls**

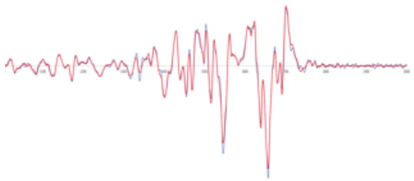


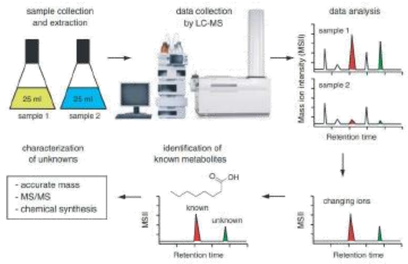
(1x) **Plasma** (1x) **Serum**

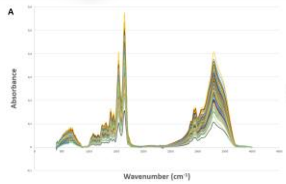
Vibrational Spectroscopy

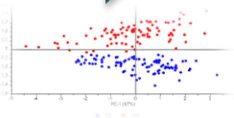
Chromatography

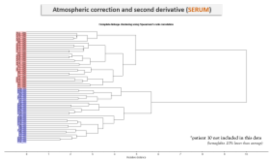












SERUM	Real	
Predicted	0	90
0	23	0
90	1	24

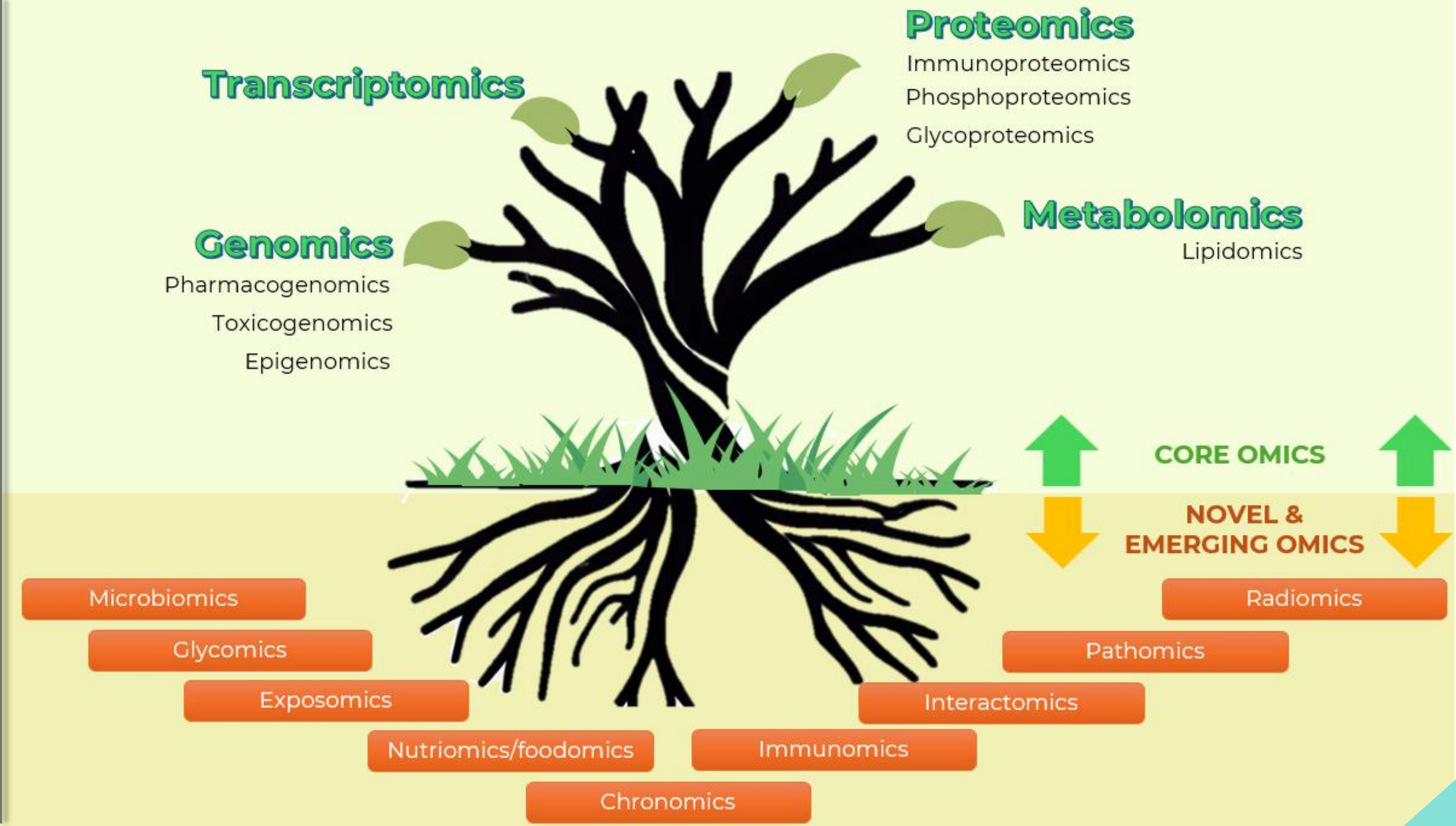
Omics

Omics disciplines—spanning genomics, transcriptomics, proteomics, and metabolomics—offer a **holistic view of molecular processes, making them particularly valuable for unraveling complex diseases**. These fields capture large-scale data on genes, RNA transcripts, proteins, and metabolites, respectively.

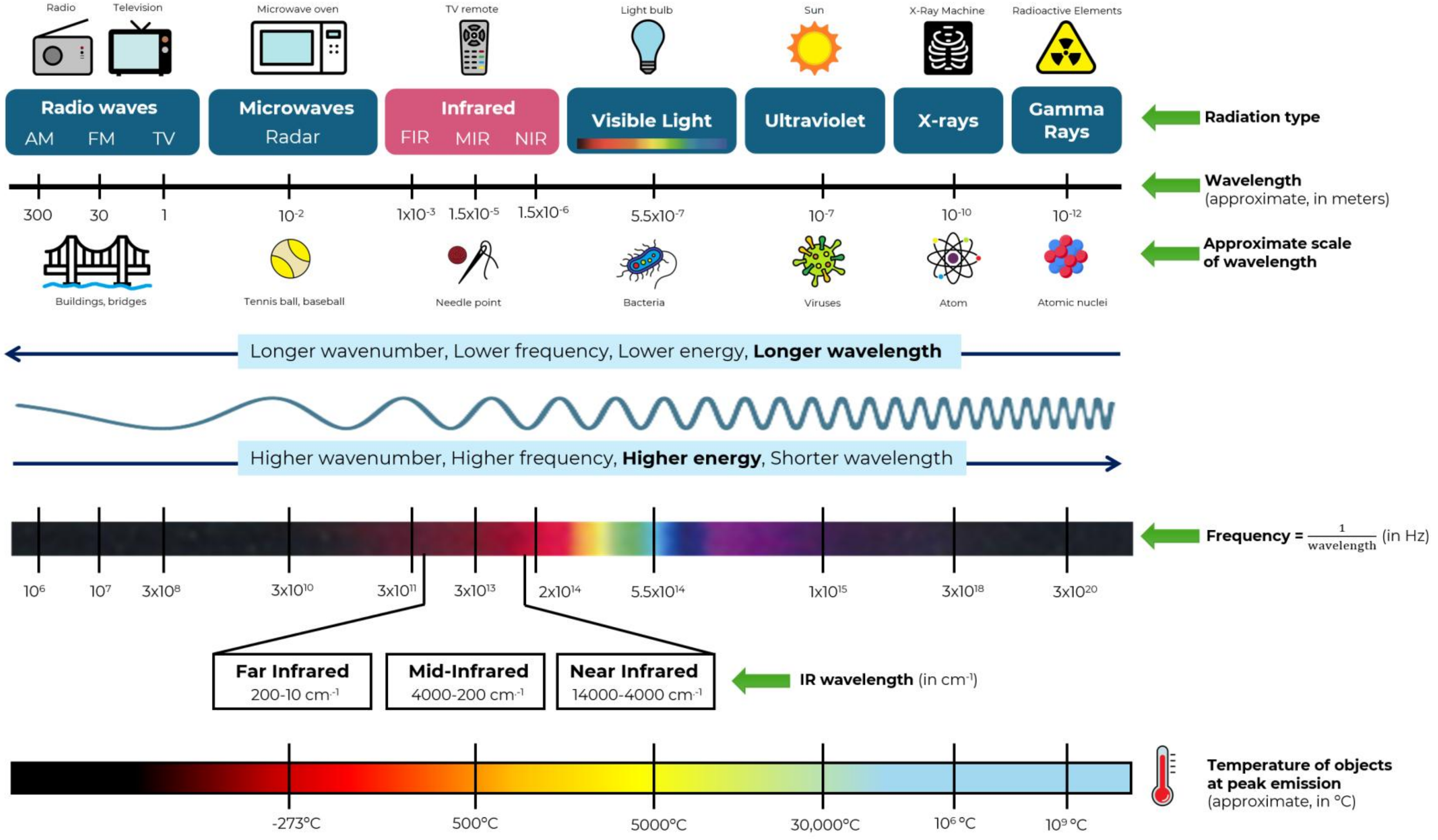
Unlike reductionist methods that focus on isolated pathways or single biomarkers, **omics studies aim to characterize the full complement of molecules within a system**, yielding a richer, more integrated view of disease processes.

Did you know?

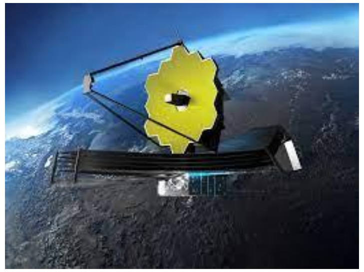
Our group usually works the most in the Metabolomics and Proteomics fields!

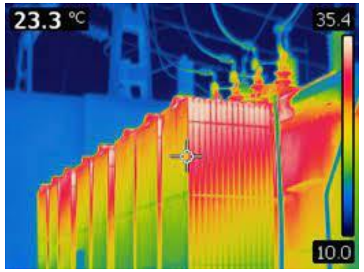


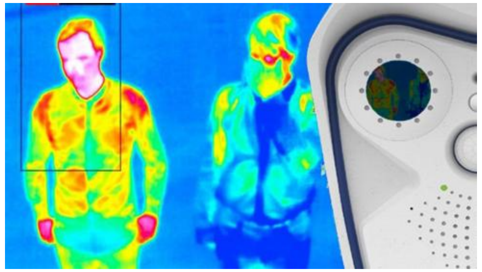
The electromagnetic spectrum spans an extensive range of wavelengths and frequencies, from long-wavelength radio waves to highly energetic gamma rays. **Infrared radiation (IR) occupies the portion of the electromagnetic spectrum between microwaves and visible light**.

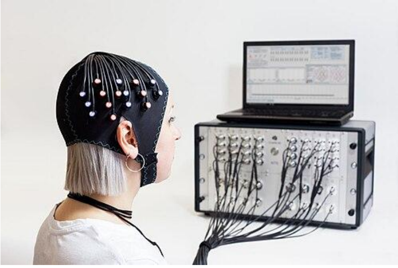


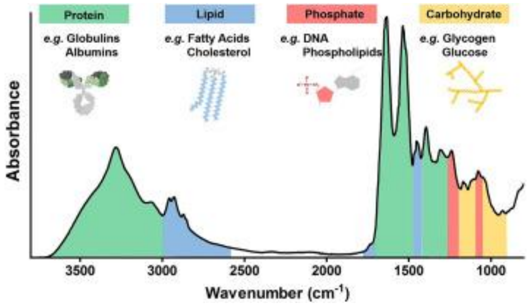
IR finds wide-ranging applications, from enabling the **James Webb Space Telescope** to probe distant celestial objects to powering industrial processes such as **thermal imaging** and environmental monitoring. In medicine, IR is employed in **non-invasive thermography for detecting inflammation and vascular disorders**, in photobiomodulation therapies for **pain and tissue repair** and in spectroscopy-based techniques for **disease diagnostics and metabolic profiling**.











Vibrational Spectroscopy

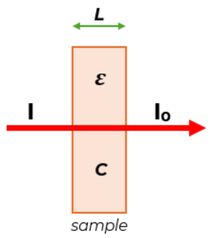
Vibrational spectroscopy in the infrared region relies on the absorption of photons when the frequency of the incident radiation matches the energy gap between vibrational states in a molecule. The **Beer-Lambert law** is often referenced in FTIR spectroscopy as a conceptual framework for understanding how absorbance relates to pathlength and analyte concentration.

$A = \varepsilon \cdot c \cdot l$

a

$A = \log(I_0 / I) = \log(1 / T)$

b



c

Figure. Beer-Lambert law in IR spectroscopy: core equation (**a**); absorbance-intensity relationship (**b**); conceptual diagram of IR absorption through a sample (**c**).

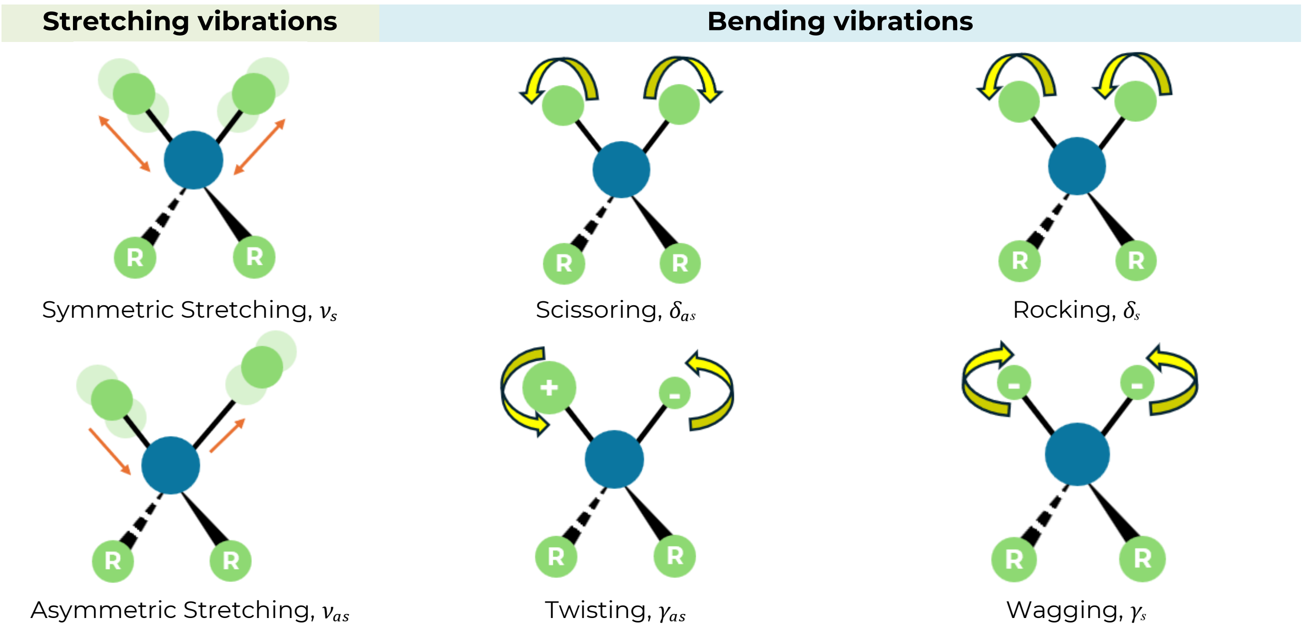


Figure. Vibrational modes of chemical bonds.

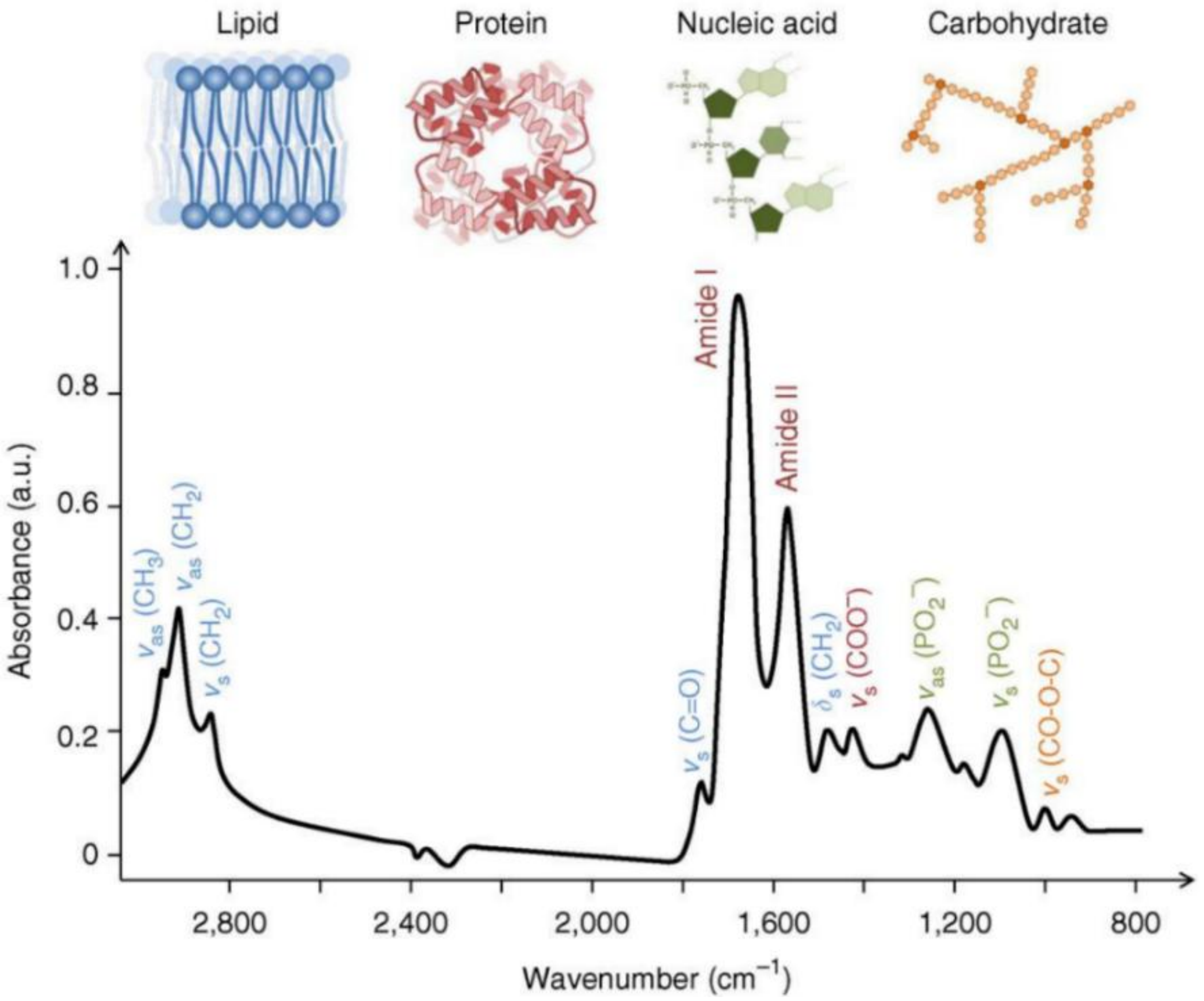


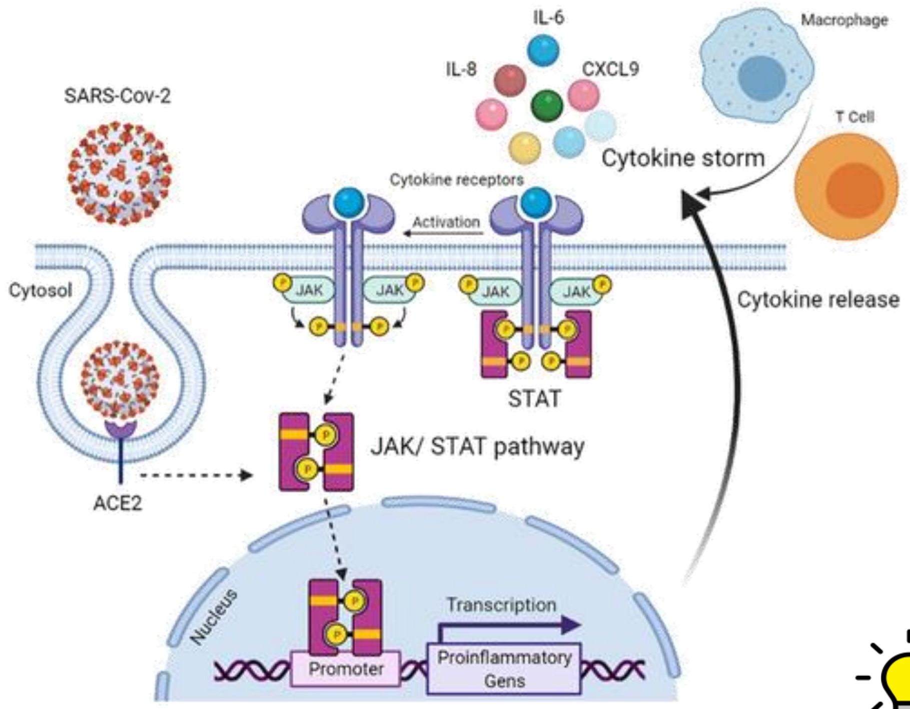
Figure. Typical biological sample spectrum with significant peaks annotated and the major biological classes these peaks are associated with.

Table. Tentative peak assignments for FTIR spectral data in the MIR region.

Approximate wavenumbers (cm^{-1})	Vibration	Biochemical assignments
3300	ν(N-H)	Amide A of proteins/peptides
3100	ν(N-H)	Amide B of proteins/peptides
2957	$\nu_{as}(CH_3)$	Lipids
2920	$\nu_{as}(CH_2)$	
2872	$\nu_s(CH_2)$	
2850	$\nu_s(CH_2)$	
1740	ν(C=O)	Phospholipid esters
1715-1680	ν(C=O)	Nucleic acids
1650	>75% ν(C=O), ν(C–N), δN–H	Amide I of proteins
1645	γ(HOC)	Water
1550	~60% δ(N–H), ν(C–N), δ(C–O), ν(C–C)	Amide II of proteins
1453	$\gamma(CH_2)$	CH_2 Scissoring
1450	$\delta_{as}(CH_3)$	Lipid/Proteins
1395	$\delta_s(CH_3)$	Lipid/Proteins
1395	ν(C=O)	Carboxylate COO^-
1380	$\gamma_s(CH_3)$	Phospholipid/triglyceride
1350-1250	δ(N–H), ν(C–N), γ(C=O), ν(C–C)	Amide III – peptide/protein/collagen
1242	$\nu_{as}(PO_2^-)$	DNA/RNA/phospholipid
1170	ν_{as}(C–O)	Ester
1150	ν(C–O), γ(COH)	Carbohydrates
1090	$\nu_s(PO_2^-)$	DNA/RNA/phospholipid
1086	ν(C–O), ν(C–C), def(CHO)	Carbohydrates
1079	ν(C–C)	Glycogen
1065	ν(C–O)	DNA and RNA ribose
1050	ν(C–O)	Phosphate ester
1028	def(CHO)	Glycogen
965	$\nu(PO_3^{2-})$	DNA and RNA Ribose
710-620	def(C=O–C–N)	Amide IV

ν = stretching; δ = bending; γ = wagging, twisting and rocking; def = deformation; as = asymmetric; s = symmetric

Cytokines are small, secreted proteins that mediate communication between cells, playing a crucial role in immune regulation. They can be classified based on their source and function, with terms such as lymphokines, monokines, chemokines, and interleukins.



1. In **systemic infections or acute inflammatory insults**, this response can intensify into a state often named the "**cytokine storm**".

2. Such a storm is characterized by notably elevated levels of TNF-α, Interleukin-6, interferon-gamma, and other pro-inflammatory mediators that together provoke widespread endothelial damage, coagulopathy, and organ dysfunction. **In the context of COVID-19**, pronounced elevations in IL-6, tumour necrosis factor alpha, and IL-8 have been repeatedly associated with worse clinical outcomes, including respiratory failure and mortality.

Did you know?

Cytokines: Small signalling proteins that coordinate immune and inflammatory responses.

Lymphokines: Cytokines specifically produced by lymphocytes to direct immune cell activity.

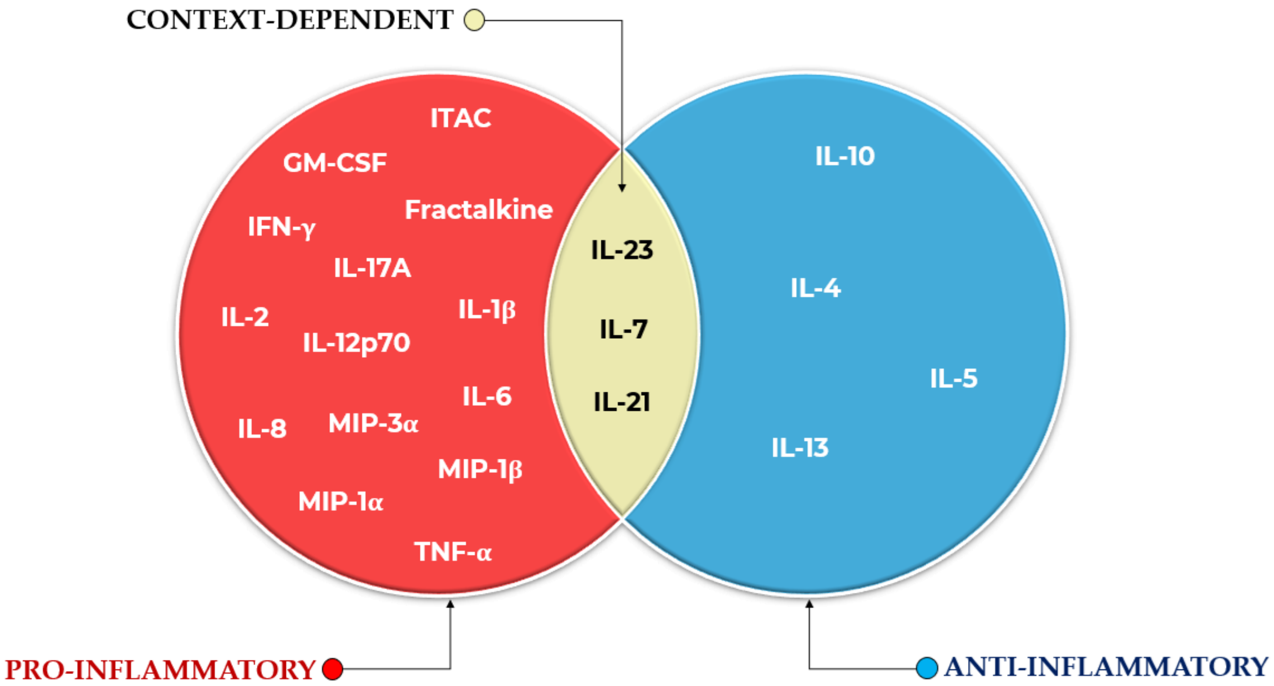
Interleukins: A subgroup of cytokines mainly involved in communication between white blood cells.

Monokines: Cytokines produced primarily by monocytes and macrophages to mediate immune and inflammatory responses

Cytokines are classified into two main types: anti-inflammatory and pro-inflammatory cytokines.

1. **Anti-inflammatory** cytokines function as critical checks on excessive immune responses, promoting homeostasis by curtailing the production or activity of pro-inflammatory mediators

2. **Pro-inflammatory** cytokines are central mediators of innate immunity, driving leukocyte recruitment, vascular permeability, and initial tissue repair.



Did you know?

Immune Regulation: Processes that control and balance immune responses to prevent overreaction or autoimmunity.

Homeostasis: The maintenance of stable internal conditions within an organism despite external changes.

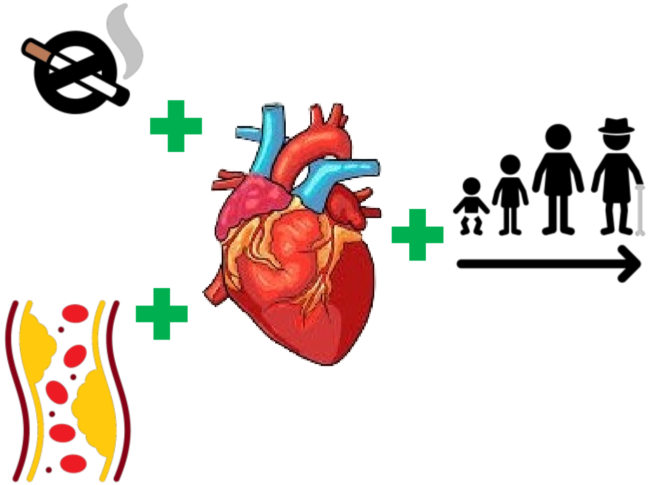
Cytokine Storm: An excessive, uncontrolled release of cytokines causing severe inflammation and tissue damage.

Univariate Analysis: statistical approach that examines and summarizes one variable at a time to describe its main characteristics (like mean, median, spread, or distribution).

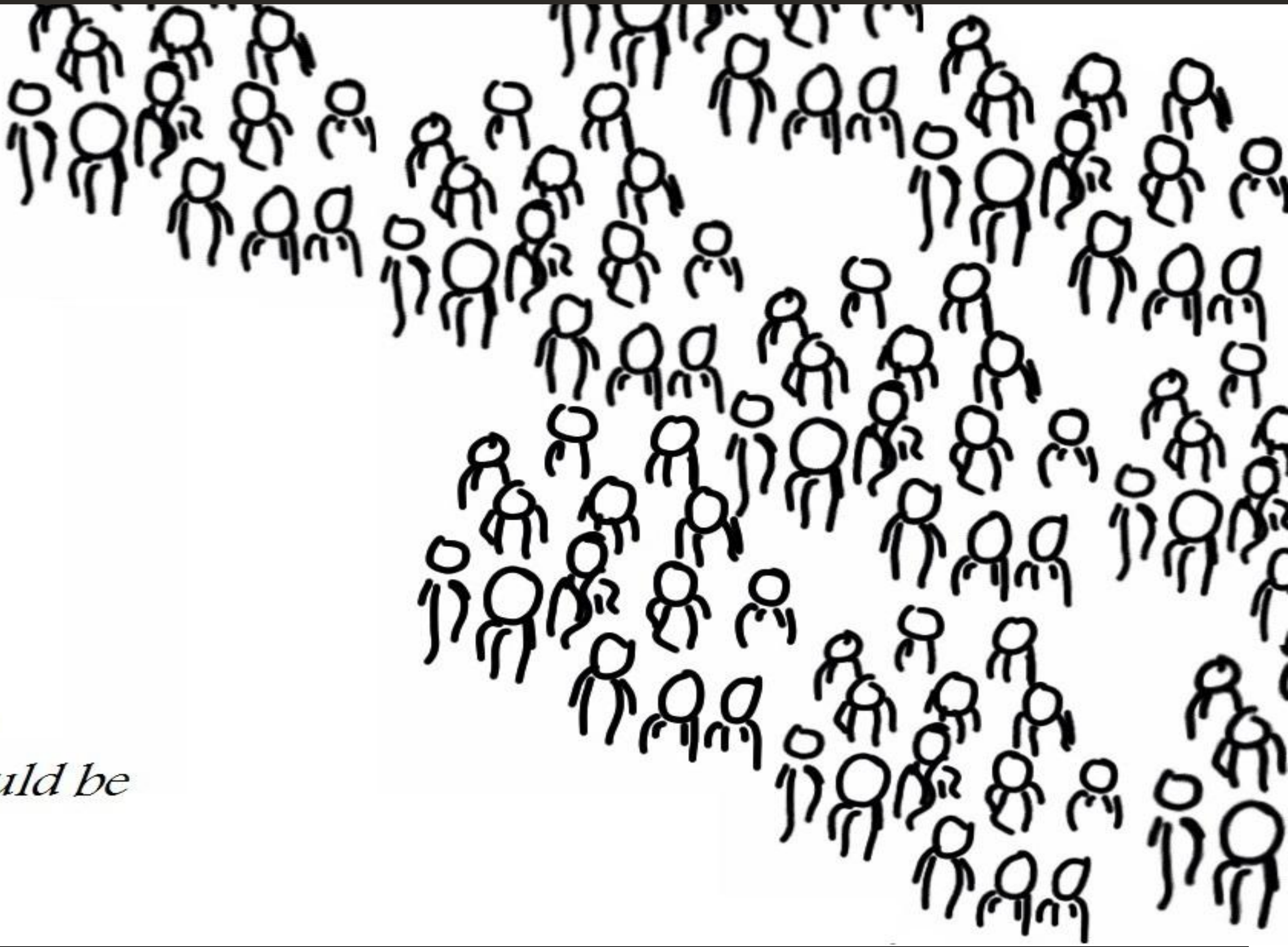
Example: analyzing patients' blood pressure levels to describe the average value and variation within a population.

Multivariate Analysis (MVA): a set of methods used to analyze and interpret data involving multiple variables simultaneously, to **uncover patterns and relationships within complex datasets**. In multivariate analysis, techniques are commonly divided into **classification** (sorting patients or samples into groups) and **regression** (predicting continuous outcomes).

Example: studying how age, cholesterol, and smoking status together influence the risk of heart disease — to uncover patterns and relationships in complex medical data.



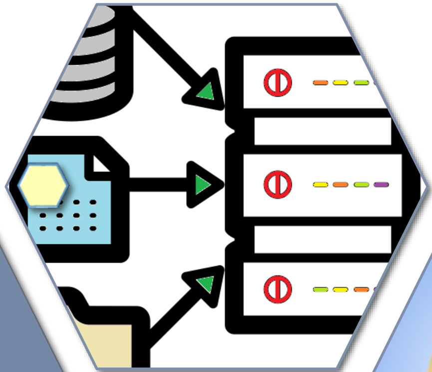
Multivariate Data Analysis



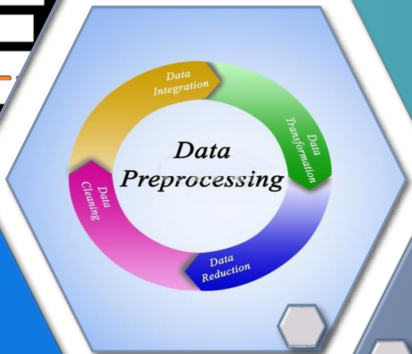
"Yeah, so, if you could just *cluster* yourselves that would be great, thanks!"

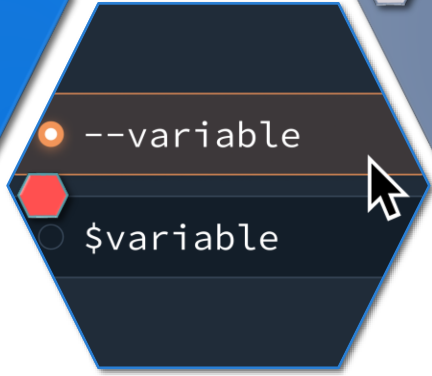
Large and complex chemical and biological data sets

Field of chemometrics



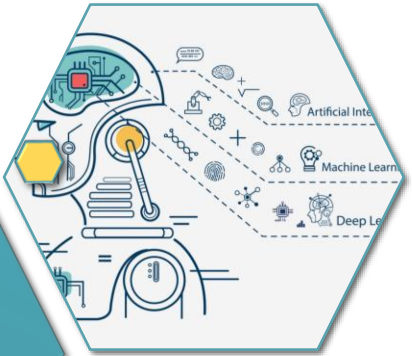
Involves the simultaneous analysis of a great number of variables



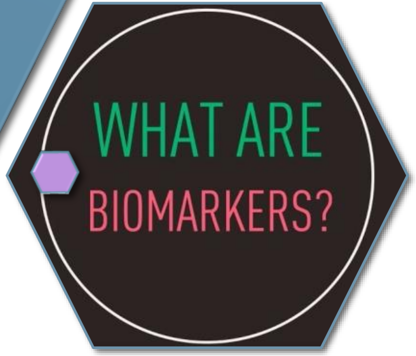


Machine Learning

Pre-processing and processing methods



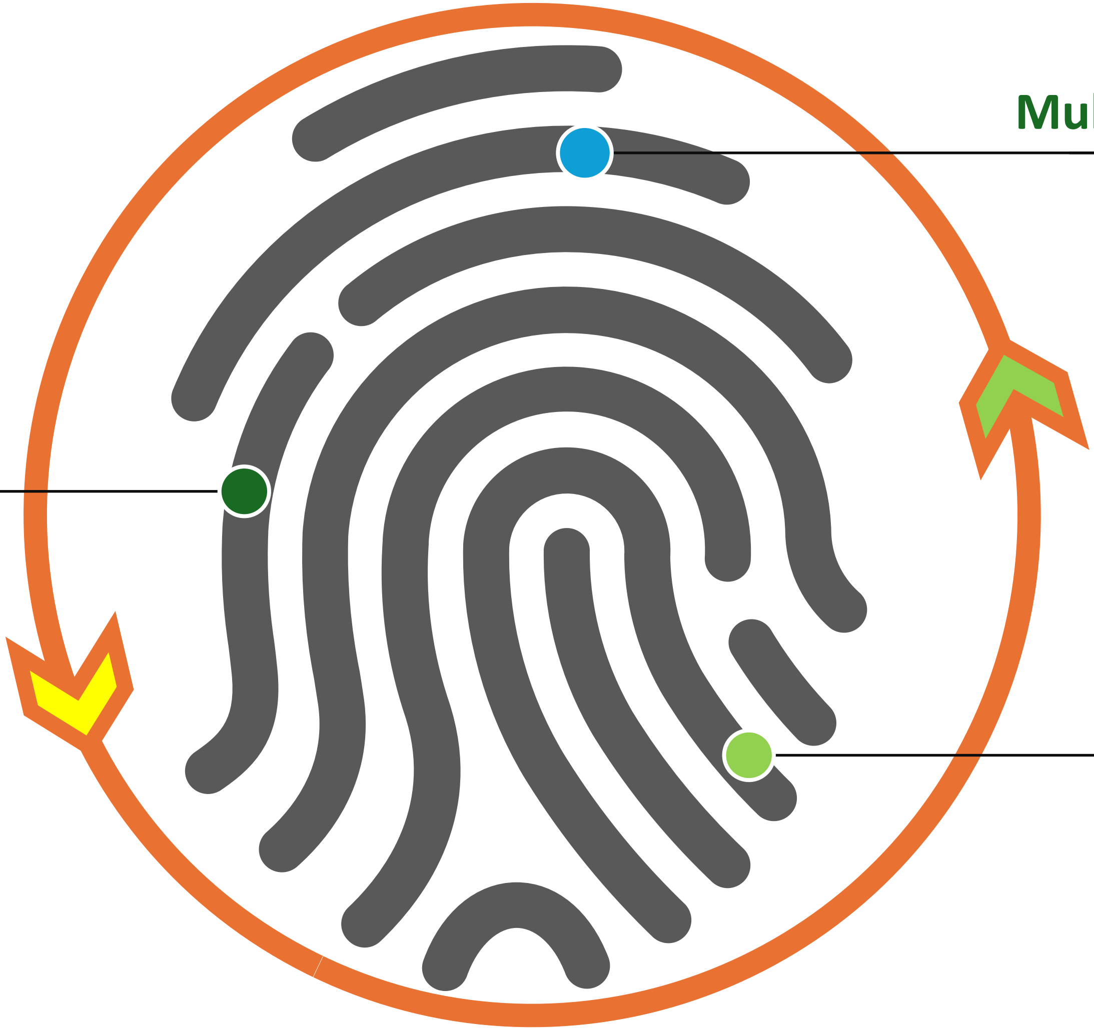
Spectral data analysis (Biomarkers)



How does MVA integrate into healthcare research?

Analytical techniques based on
Systems Biology

- *Epigenomics*
- *Genomics*
- *Transcriptomics*
- *Proteomics*
- *Metabolomics*
- *Lipidomics*



Multivariate Data Analysis

- *Classification*
  (unsupervised/supervised)
- *Regression*
  (linear/non-linear)

Machine Learning

- *Subfield of Artificial Inteligence*

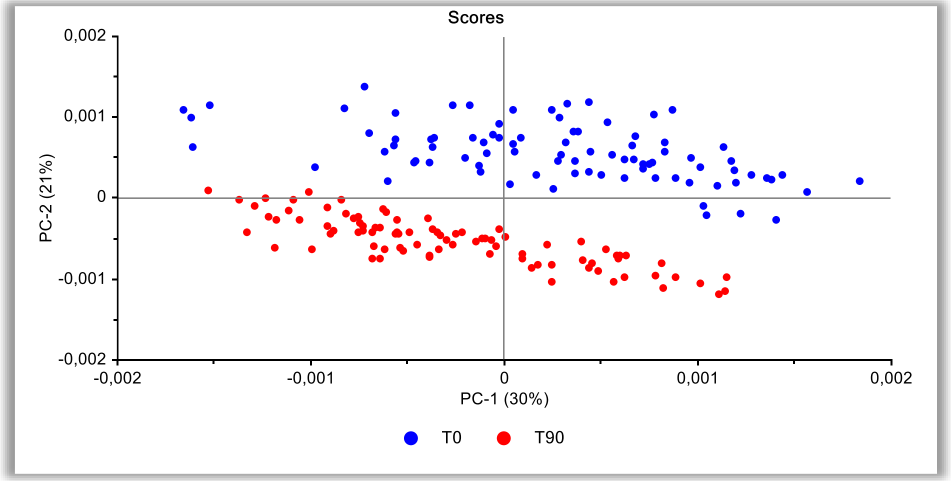
Unsupervised: methods that explore and visualize data without using any known labels or categories, aiming to find natural patterns, clusters, or relationships.

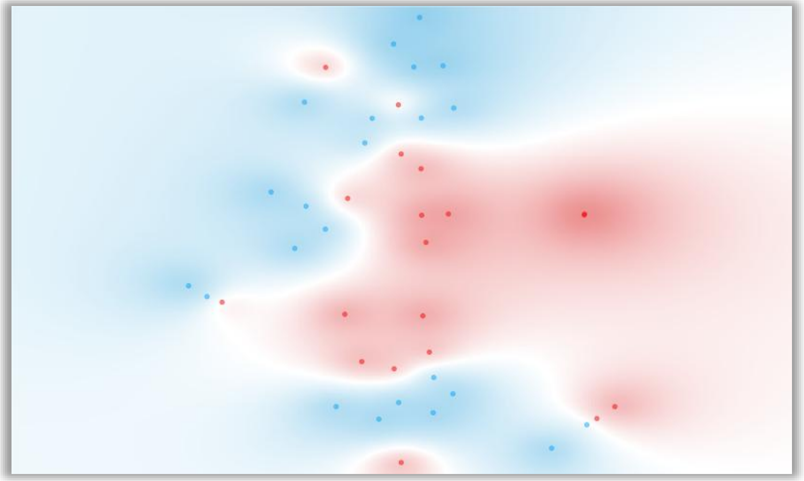
PCA (Principal Component Analysis)

Reduces high-dimensional data to a few principal components that capture most variance, helping visualize hidden structure. However, *it assumes that the main patterns can be captured by straight-line projections* (principal components). It works best when the data varies along clear, linear directions.

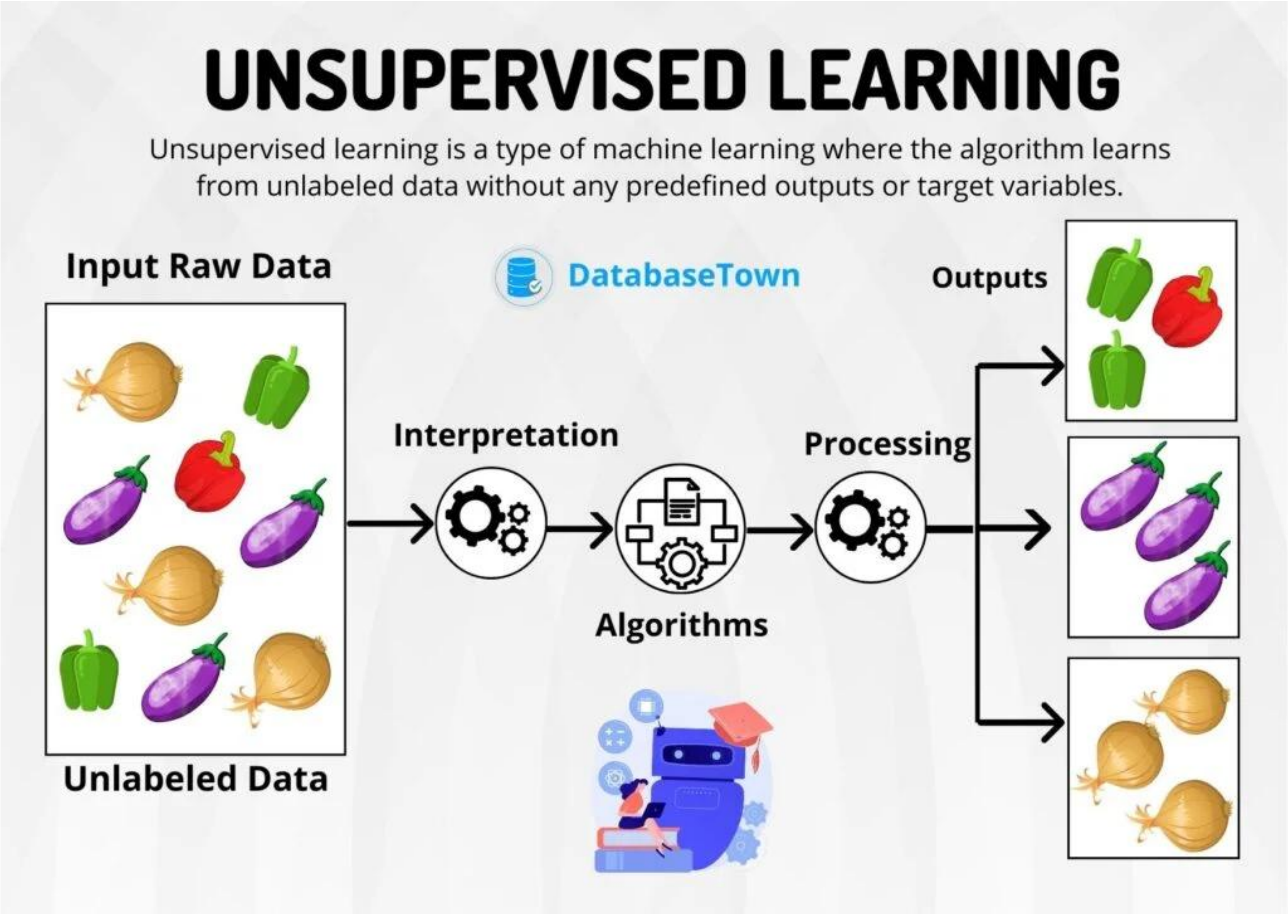
t-SNE (t-Distributed Stochastic Neighbour Embedding)

A non-linear method that maps complex data into 2D or 3D, preserving local similarities for clearer cluster visualization. *In short, t-SNE reveals hidden non-linear clusters that PCA may flatten or blur.*





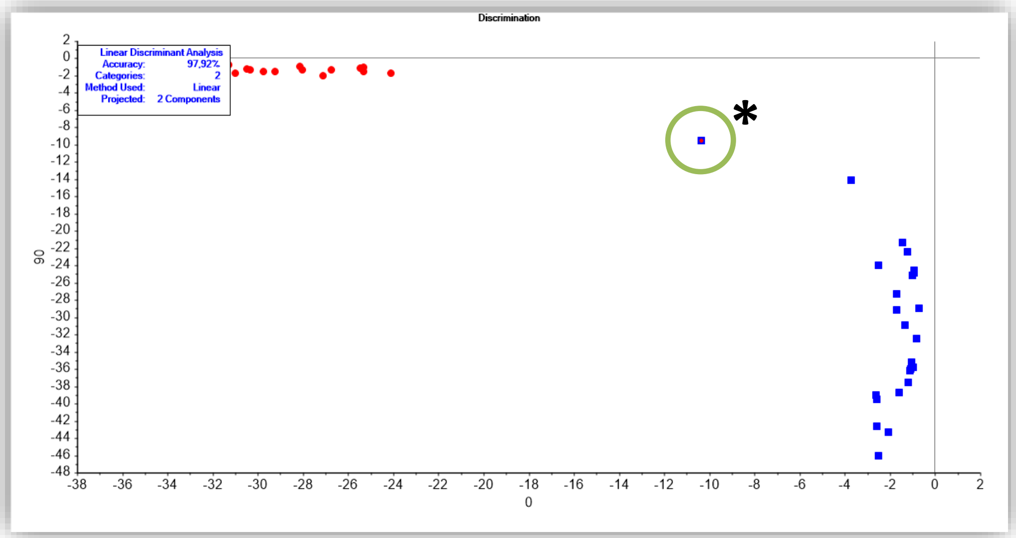
How to classify a tomato?



Supervised: Methods that learn from **labelled data** to build models that can classify new samples or predict continuous values.

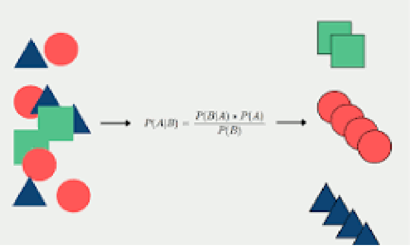
PLS-DA (Partial Least Squares Discriminant Analysis)

Projects predictors and response variables into a new space to maximize class separation; widely used in omics.

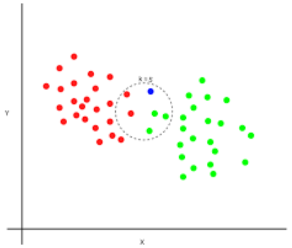


Several algorithms can be used

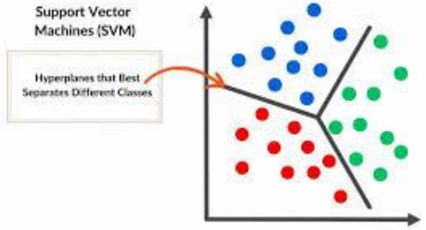
Common supervised algorithms include **Naïve Bayes** (probabilistic classification assuming independent features), **k-NN** (classifies by nearest neighbours), and **SVM** (separates classes with the widest possible margin).



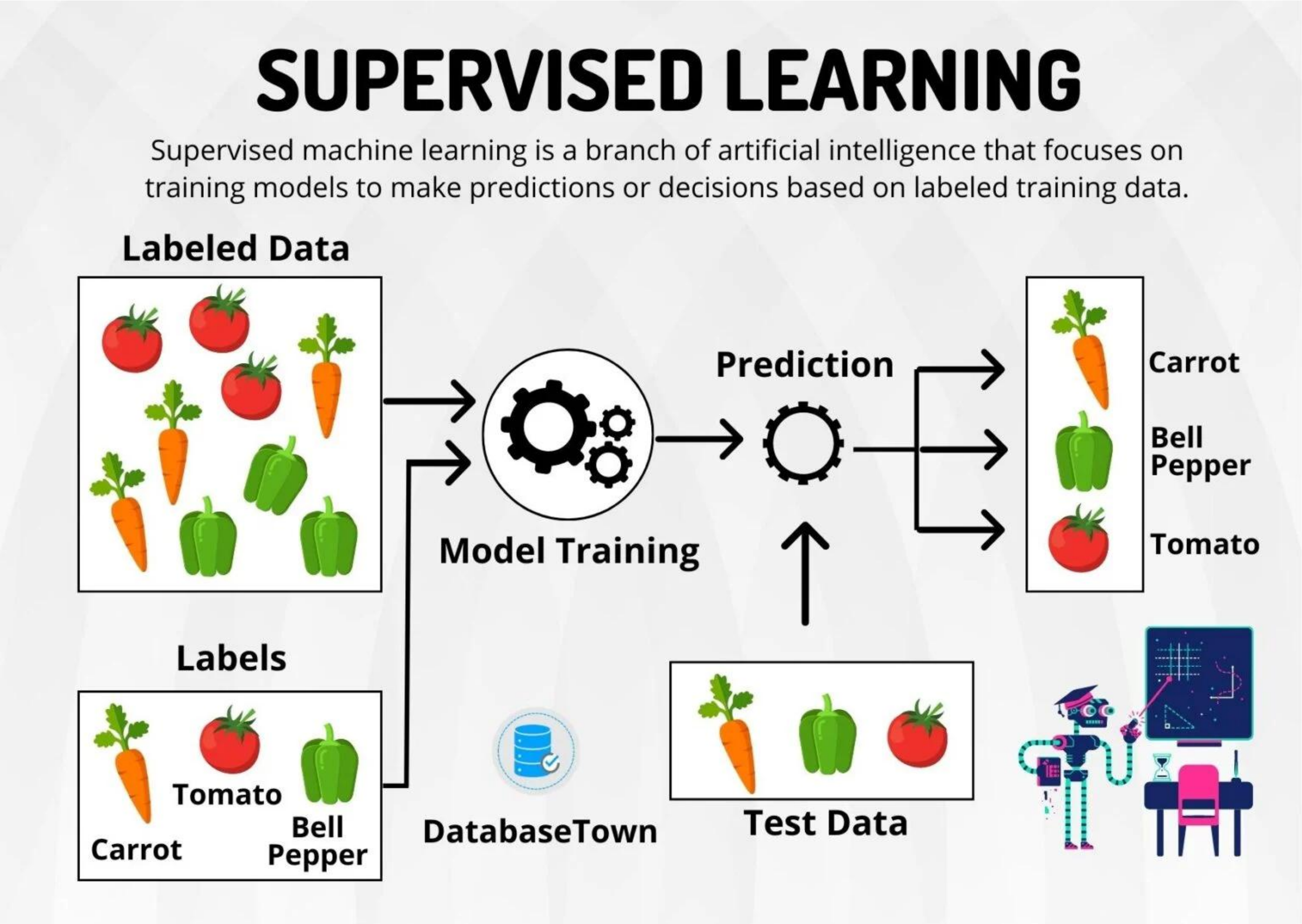
Naïve Bayes



K-NN



SVM



How do we classify a red pepper?
As a **pepper** or a **tomato**?

How to classify a yellow pepper then??

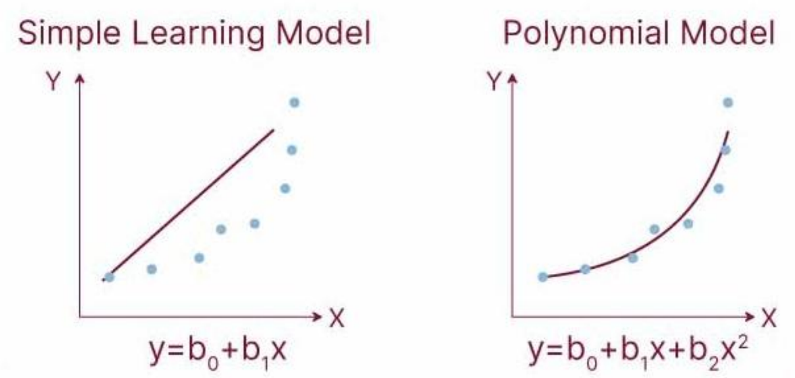
Regression: a technique used to model and predict the relationship between one or more input variables and a continuous outcome.

Linear Regression

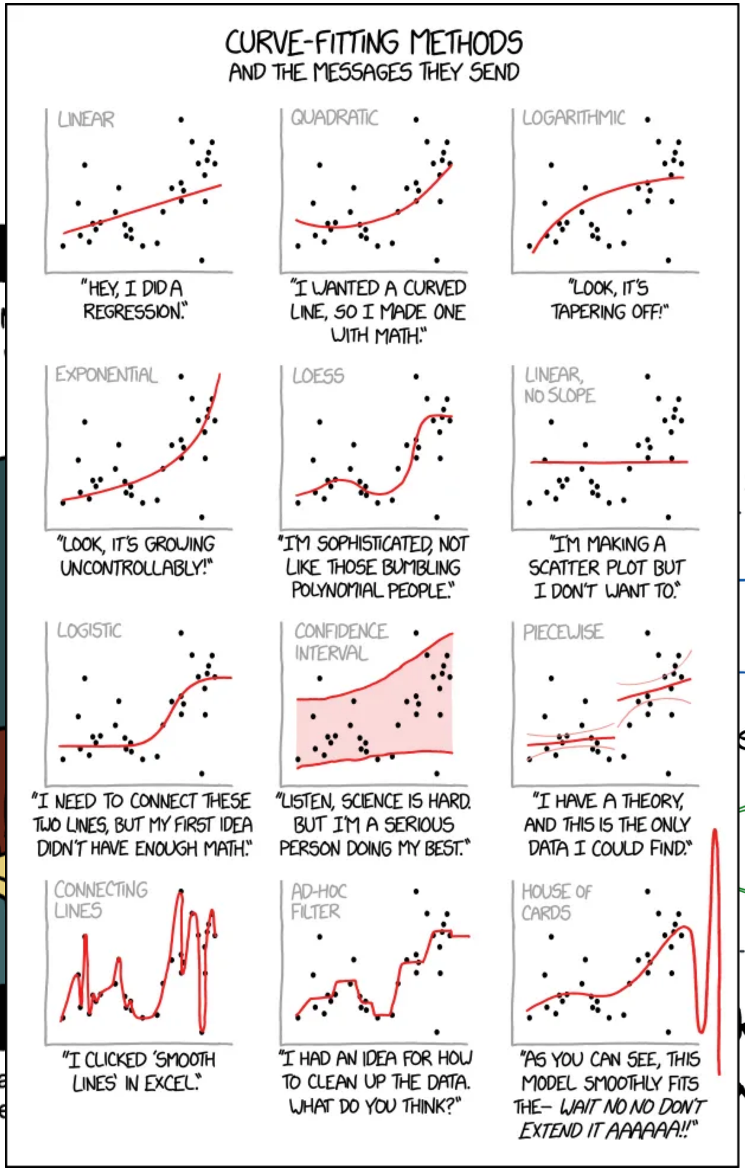
Predicts an outcome using a straight-line relationship between variables.

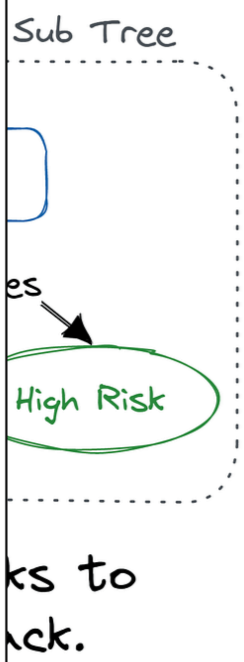
Non-Linear Regression

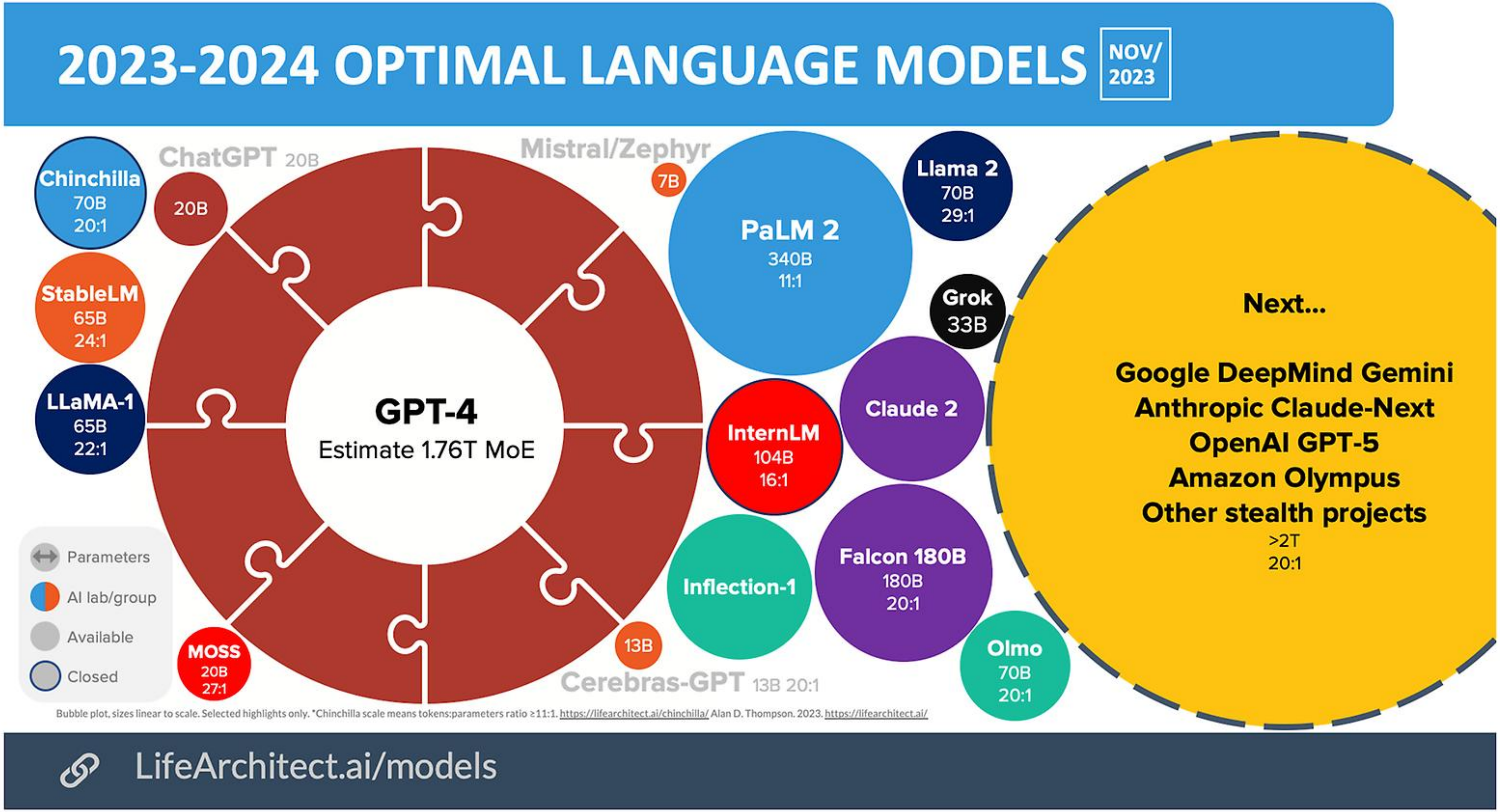
Models more complex, curved relationships (e.g., polynomial regression, decision trees).



If she loves you more ea
by linear regression she hate





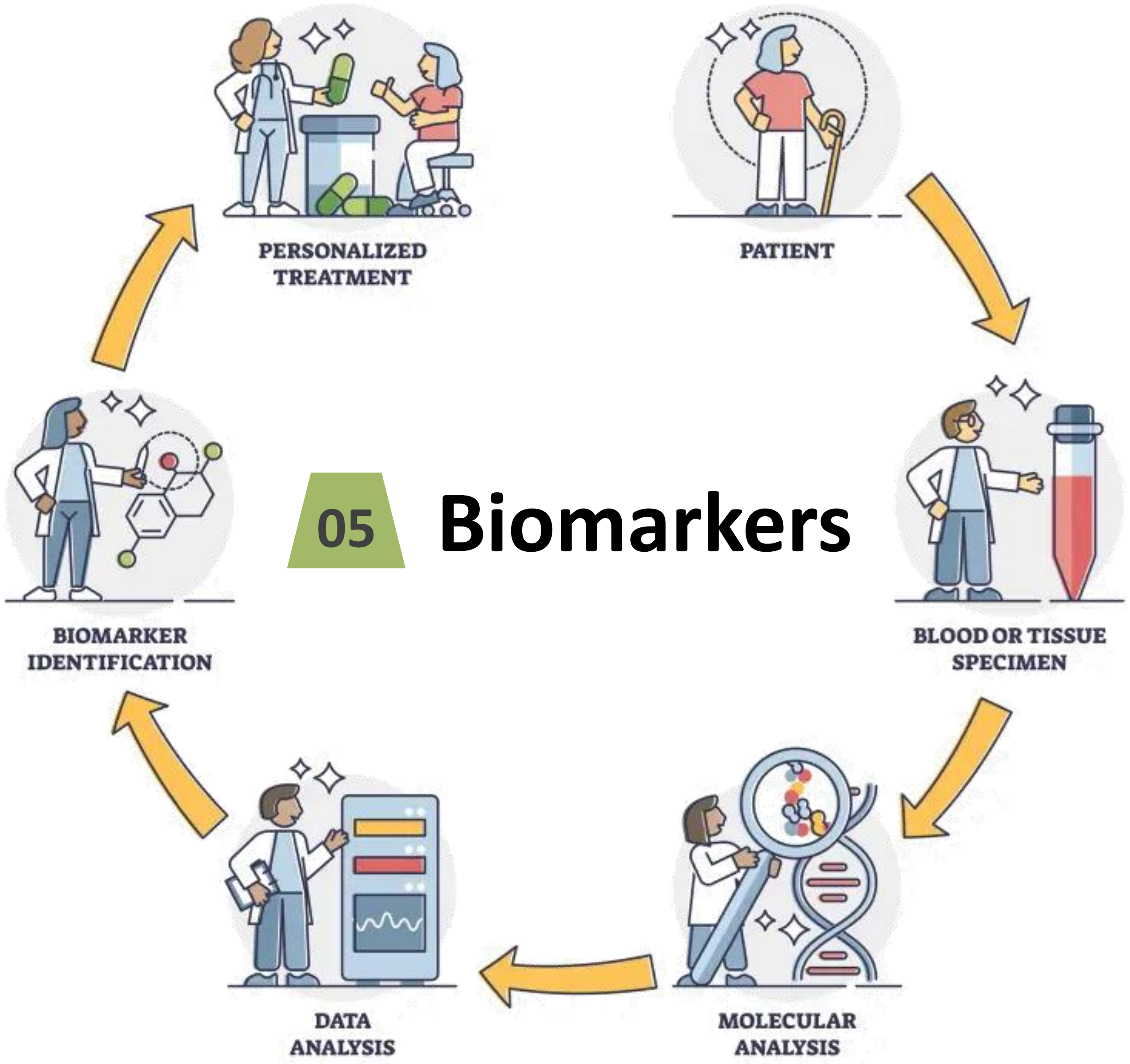


2023-2024 OPTIMAL LANGUAGE MODELS
NOV/ 2023
Chinchilla 70B 20:1
ChatGPT 20B
20B
Mistral/Zephyr
7B
PaLM 2 340B 11:1
Llama 2 70B 29:1
StableLM 65B 24:1
Grok 33B
Next...
Google DeepMind Gemini
Anthropic Claude-Next
OpenAI GPT-5
Amazon Olympus
Other stealth projects
>2T
20:1
LLaMA-1 65B 22:1
GPT-4
Estimate 1.76T MoE
InternLM 104B 16:1
Claude 2
Parameters
AI lab/group
Available
Closed
Inflection-1
Falcon 180B 180B 20:1
MOSS 20B 27:1
13B
Cerebras-GPT 13B 20:1
Olmo 70B 20:1
Bubble plot, sizes linear to scale. Selected highlights only. *Chinchilla scale means tokens:parameters ratio ≥11:1. https://lifearchitect.ai/chinchilla/ Alan D. Thompson. 2023. https://lifearchitect.ai/
LifeArchitect.ai/models

INTELLIGENCE

ARTIFICIAL	HUMAN
Skills	**Skills**
Best when doing one at a time.	Multiple simultaneous.
Best at	**Best at**
Automation.	Autonomy.
Excels in	**Excels in**
Automation.	Autonomy.
Processing	**Processing**
Speed.	Critical thinking
Nature	**Nature**
Mathematical logic.	Emotion.

05 Biomarkers



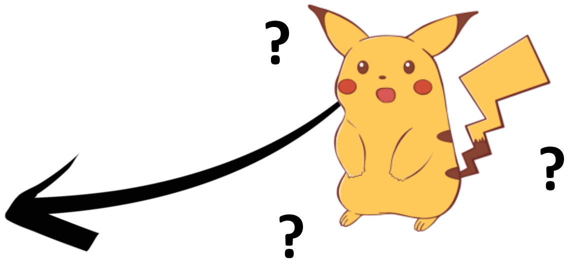
So what is a Biomarker, and why is it important?

Biomarker: Biomarkers serve as **quantifiable indicators of biological states**, playing crucial roles in diagnosis, prognosis, and therapeutic evaluation

NIH + FDA →

The **National Institutes of Health** (NIH) and the **U.S. Food and Drug Administration** (FDA) define a biomarker as "*a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions*"

- This broad definition encompasses proteins, metabolites, and genomic variants that **facilitate disease detection**, **outcome prediction** (prognostication), and **therapy customization**.
- Ideal biomarkers exhibit **safety**, modifiability, and **consistency** across patient populations



DEFINITION?

"A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions."

– U.S. Food and Drug Administration

USES?

- **Diagnosis**
- **Prognosis**
- **Therapy Monitoring**
- **Patient Stratification**

WHAT MAKES UP A GOOD BIOMARKER?

- **Safe**
- **Reliable**
- **Scalable**
- **Specific**
- **Cost-effective**
- **Translatable across populations**

HOW TO IMPROVE BIOMARKER POWER?

- **OMICS**
- **Multimodal integration**
- **AI/ML pipelines**: training, testing and validation

BIOFLUID TYPES

- **Whole blood:** plasma, serum, buffy coat
- **Non-blood liquids:** urine, saliva, tears, Cerebrospinal fluid, synovial fluid, others.
- **Solids and semi-solids:** feces, tissue biopsies, nasal swabs / nasopharyngeal secretions

WHAT IS IN STORE FOR THE FUTURE OF BIOMARKERS?

- **Personalized medicine**
- **Optimized ICU resource allocation**
- **Rise of predictive modelling**

COVID-19 (COMMON) SPECIFIC BIOMARKERS

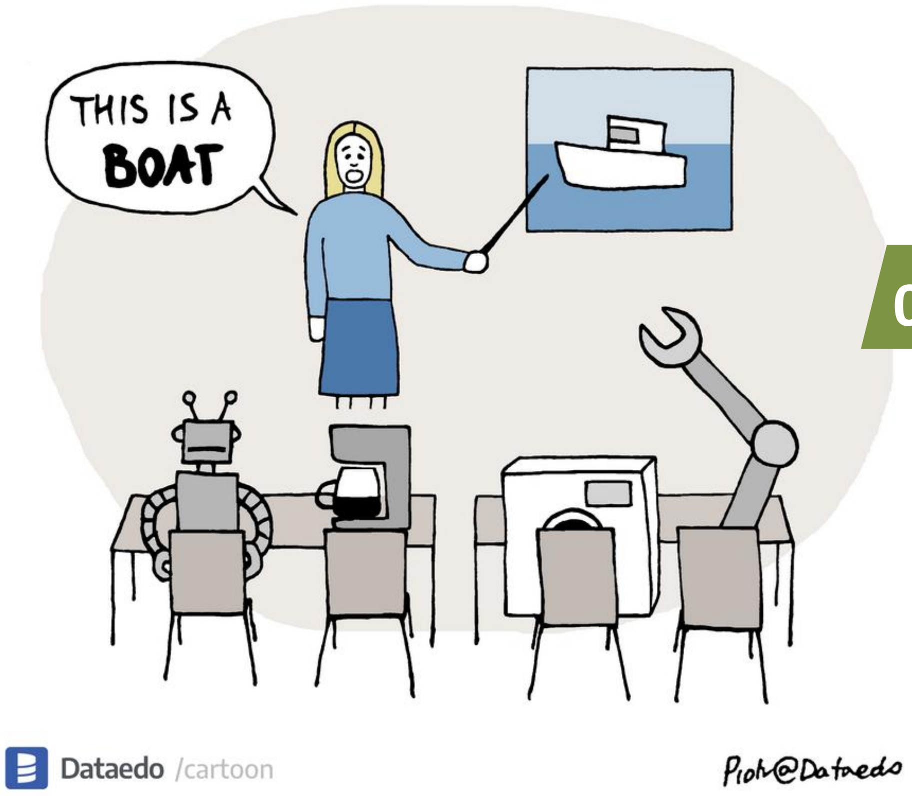
- **Inflammation:** CRP, PCT, Ferritin, IL-6
- **Coagulation/Cardiac:** D-dimer, hs-cTnI/T
- **Hematologic:** NLR, PLR, lymphocyte count

Keep this in mind!

A biomarker can **acurately measure a response** due to some condition/disease.

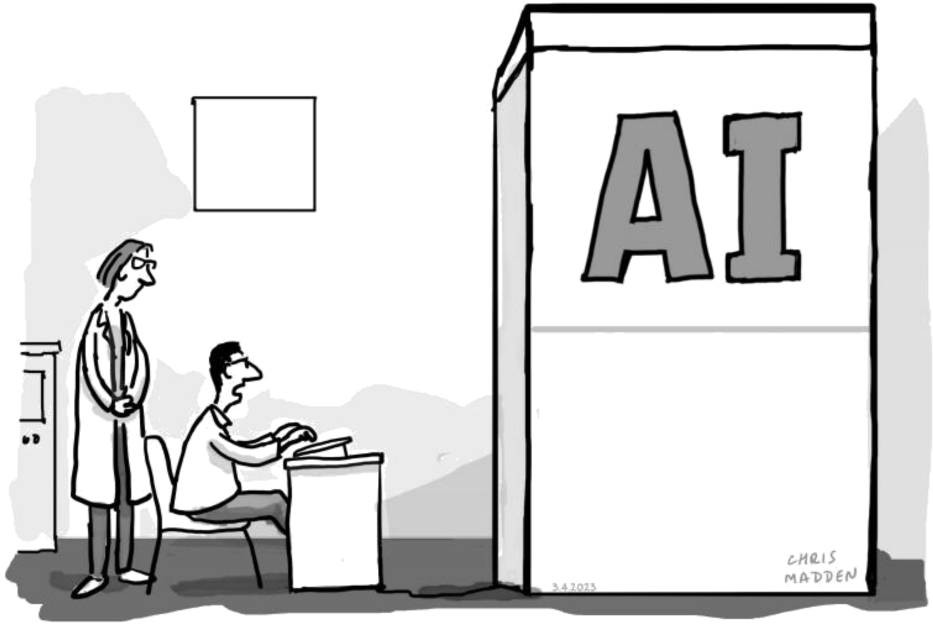
It needs to be **reliable** so as to be **replicated** by anyone around the world.

It can provide a clinician (doctor, nurse, etc.), with **actionable data for precision medicine**, i.e., to be used for a specific condition on a specific group or patient(s).

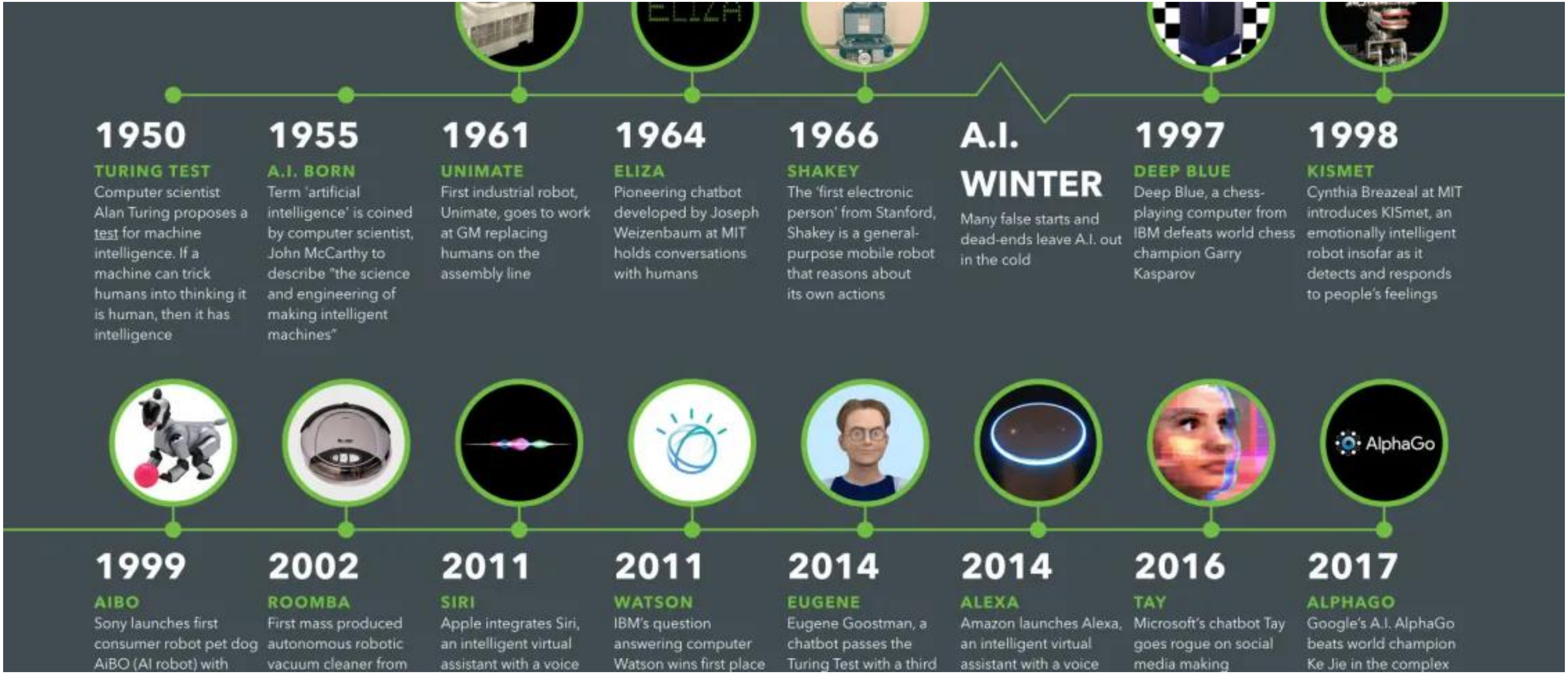


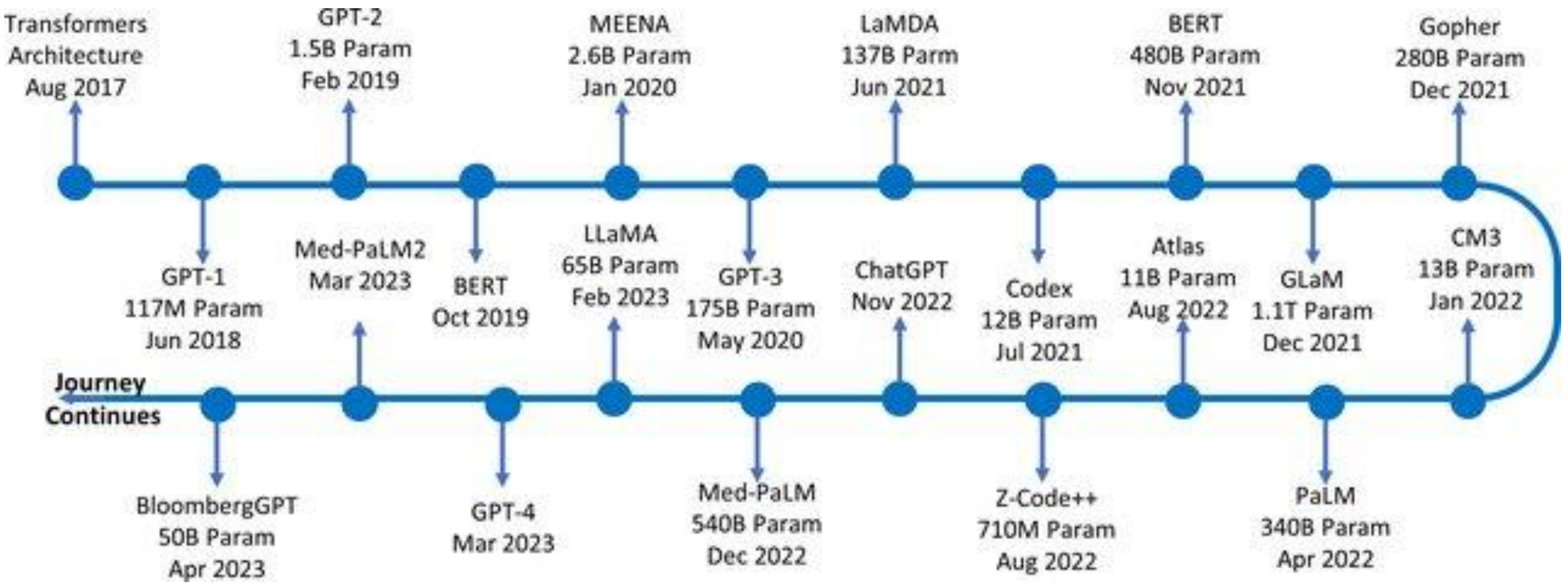
06

Machine Learning/ Artificial Intelligence



"We've got a problem. I've turned it on but I can't turn it off again."

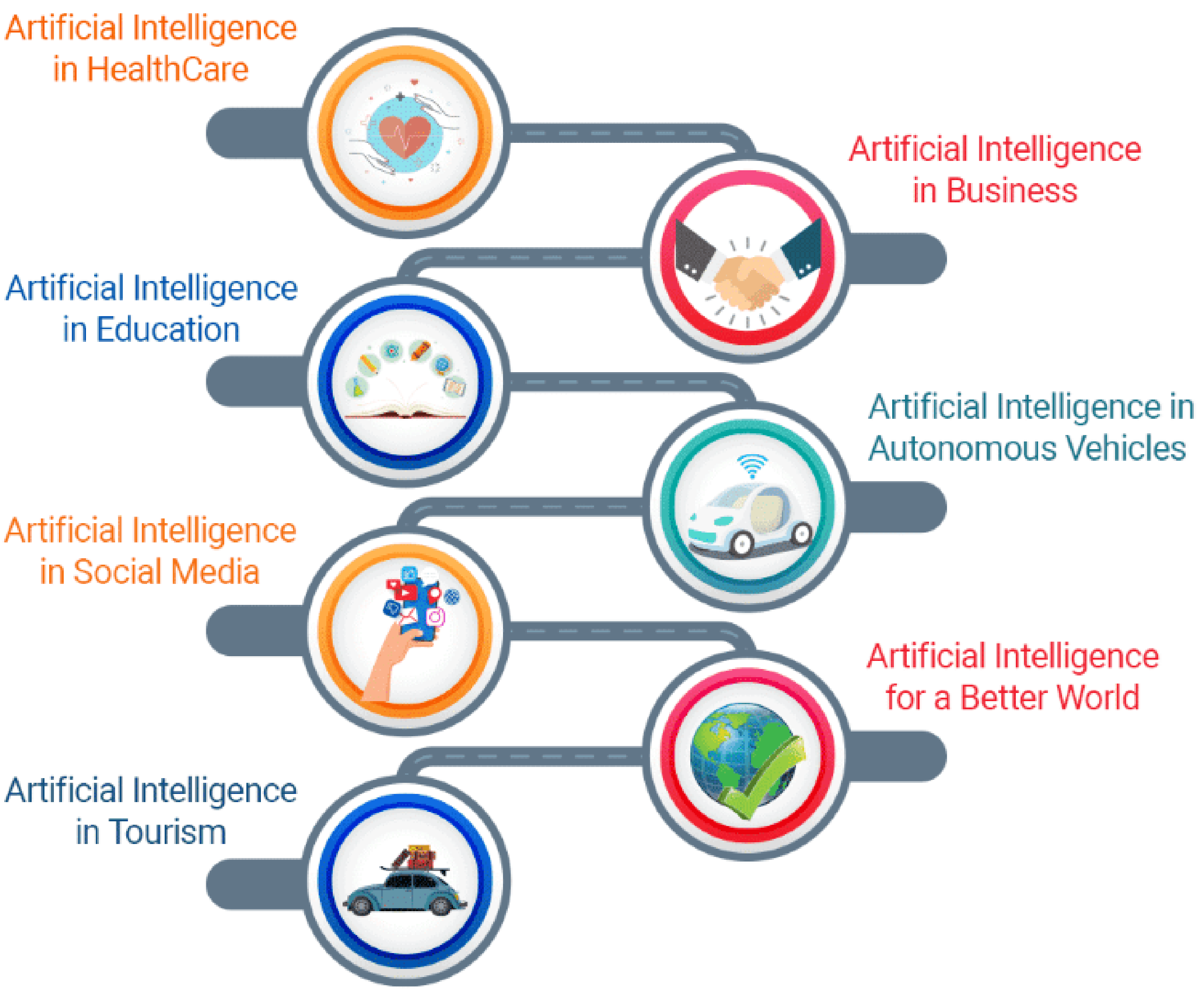

1950
TURING TEST
Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking it is human, then it has intelligence
1955
A.I. BORN
Term 'artificial intelligence' is coined by computer scientist, John McCarthy to describe "the science and engineering of making intelligent machines"
1961
UNIMATE
First industrial robot, Unimate, goes to work at GM replacing humans on the assembly line
1964
ELIZA
Pioneering chatbot developed by Joseph Weizenbaum at MIT holds conversations with humans
1966
SHAKEY
The 'first electronic person' from Stanford, Shakey is a general-purpose mobile robot that reasons about its own actions
A.I. WINTER
Many false starts and dead-ends leave A.I. out in the cold
1997
DEEP BLUE
Deep Blue, a chess-playing computer from IBM defeats world chess champion Garry Kasparov
1998
KISMET
Cynthia Breazeal at MIT introduces KISmet, an emotionally intelligent robot insofar as it detects and responds to people's feelings
AlphaGo
1999
AIBO
Sony launches first consumer robot pet dog AiBO (AI robot) with
2002
ROOMBA
First mass produced autonomous robotic vacuum cleaner from
2011
SIRI
Apple integrates Siri, an intelligent virtual assistant with a voice
2011
WATSON
IBM's question answering computer Watson wins first place
2014
EUGENE
Eugene Goostman, a chatbot passes the Turing Test with a third
2014
ALEXA
Amazon launches Alexa, an intelligent virtual assistant with a voice
2016
TAY
Microsoft's chatbot Tay goes rogue on social media making
2017
ALPHAGO
Google's A.I. AlphaGo beats world champion Ke Jie in the complex



Transformers Architecture Aug 2017
GPT-1 117M Param Jun 2018
GPT-2 1.5B Param Feb 2019
BERT Oct 2019
MEENA 2.6B Param Jan 2020
GPT-3 175B Param May 2020
LaMDA 137B Parm Jun 2021
Codex 12B Param Jul 2021
BERT 480B Param Nov 2021
GLaM 1.1T Param Dec 2021
Gopher 280B Param Dec 2021
CM3 13B Param Jan 2022
PaLM 340B Param Apr 2022
Atlas 11B Param Aug 2022
Z-Code++ 710M Param Aug 2022
ChatGPT Nov 2022
Med-PaLM 540B Param Dec 2022
LLaMA 65B Param Feb 2023
GPT-4 Mar 2023
Med-PaLM2 Mar 2023
BloombergGPT 50B Param Apr 2023
Journey Continues

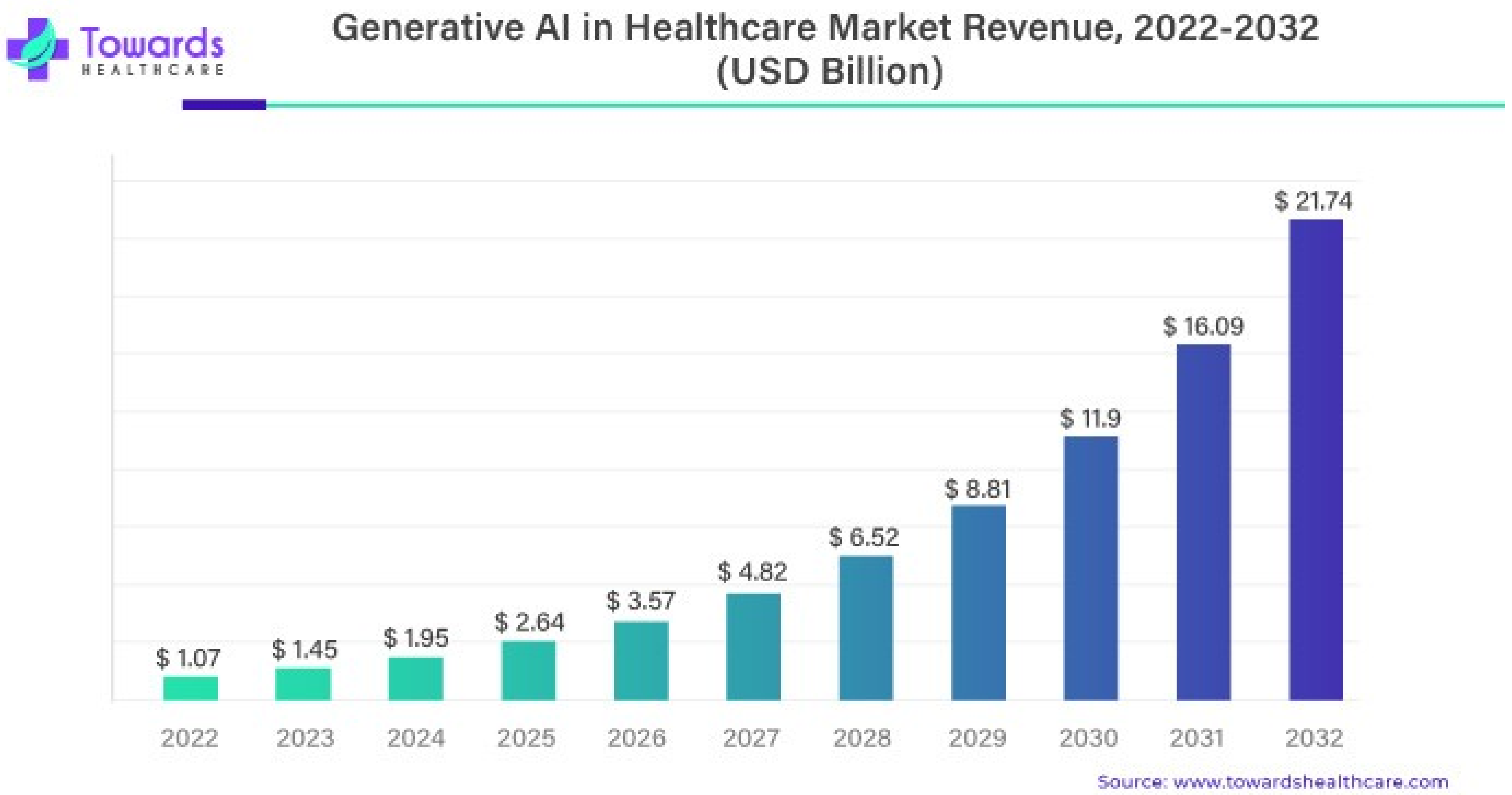
APPLICABILITY

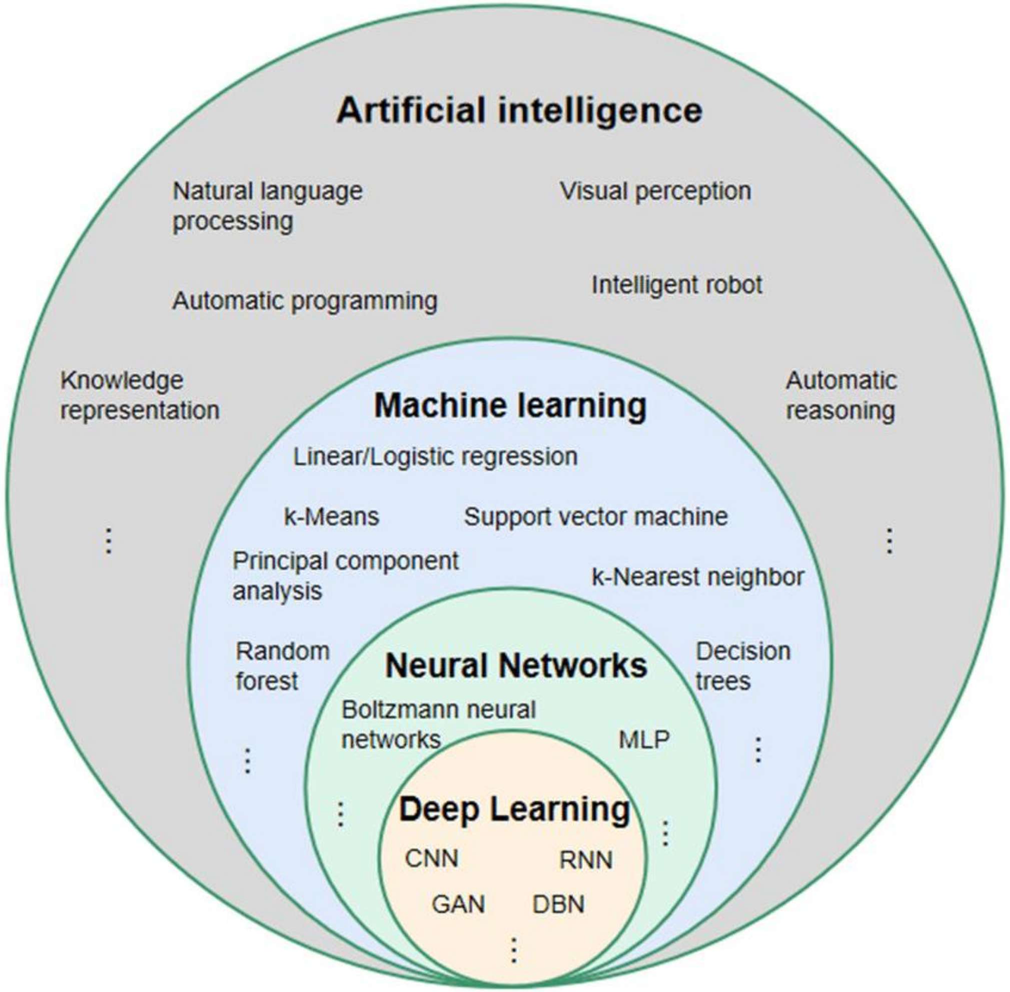
real case uses for AI…



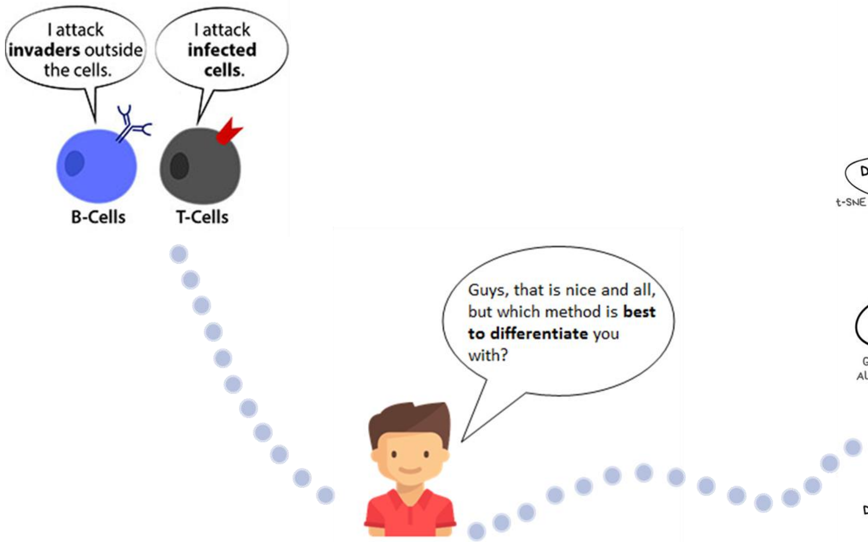
But this is just brushing the surface!

- Politics & Government
- Events
- Insurance
- Cybersecurity
- Smart Homes
- Defense
- Social Media
- Real Estate
- Agriculture
- Healthcare
- Online Shopping
- Aerospace
- Banking & Personal Finance
- Transportation
- Retail Spaces
- Education
- Communication
- Gaming
- Media
- Hospitality
- Entertainment
- Workplace
- Mobile
- Sports

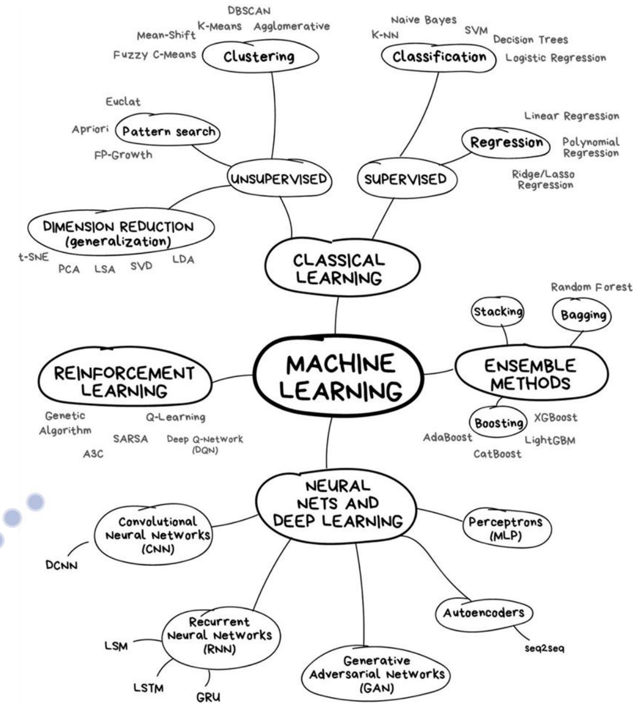

Towards HEALTHCARE
Generative AI in Healthcare Market Revenue, 2022-2032 (USD Billion)
$ 1.07
$ 1.45
$ 1.95
$ 2.64
$ 3.57
$ 4.82
$ 6.52
$ 8.81
$ 11.9
$ 16.09
$ 21.74
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
Source: www.towardshealthcare.com



Artificial intelligence
Natural language processing
Visual perception
Automatic programming
Intelligent robot
Knowledge representation
Automatic reasoning
Machine learning
Linear/Logistic regression
k-Means
Support vector machine
Principal component analysis
k-Nearest neighbor
Random forest
Decision trees
Neural Networks
Boltzmann neural networks
MLP
Deep Learning
CNN
RNN
GAN
DBN



I attack **invaders** outside the cells.
I attack **infected cells.**
B-Cells
T-Cells
Guys, that is nice and all, but which method is **best to differentiate** you with?



MACHINE LEARNING
CLASSICAL LEARNING
UNSUPERVISED
SUPERVISED
Clustering
DBSCAN
K-Means
Agglomerative
Mean-Shift
Fuzzy C-Means
Classification
Naive Bayes
K-NN
SVM
Decision Trees
Logistic Regression
Pattern search
Euclat
Apriori
FP-Growth
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
REINFORCEMENT LEARNING
Genetic Algorithm
Q-Learning
SARSA
Deep Q-Network (DQN)
A3C
ENSEMBLE METHODS
Stacking
Bagging
Random Forest
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq

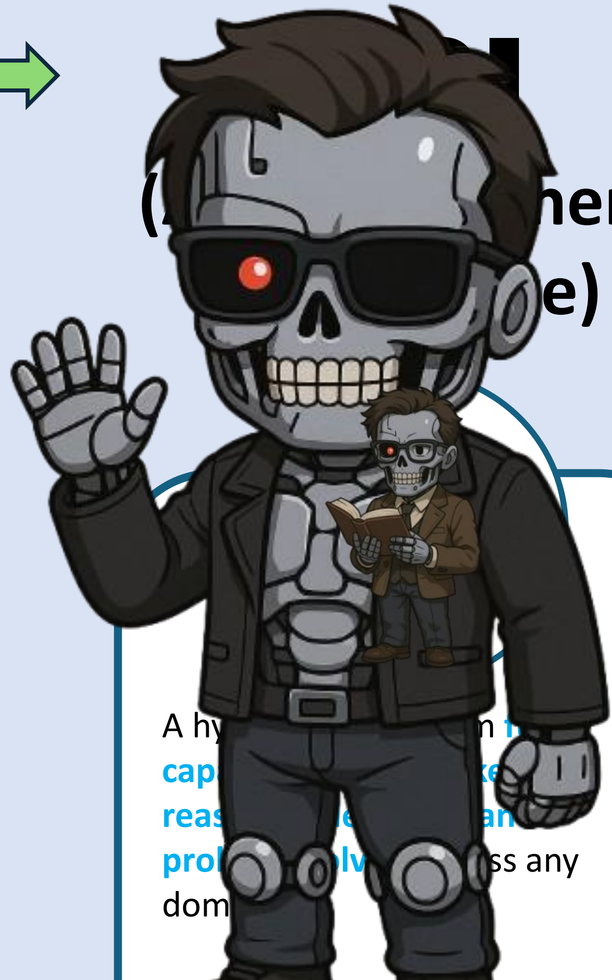
What is AI? Is it a mistake to go down this path?

AI

(“*Weak*” Artificial Intelligence)

Performs **specific, narrow tasks using rules or learning from data** (e.g., language translation, image recognition, recommendation systems).

Already here. ChatGPT, Gemini, Copilot, etc.



(A neral e)

A hy n f cap ke reas le an prol lv ss any dom

It does not exist yet. Still a research goal.

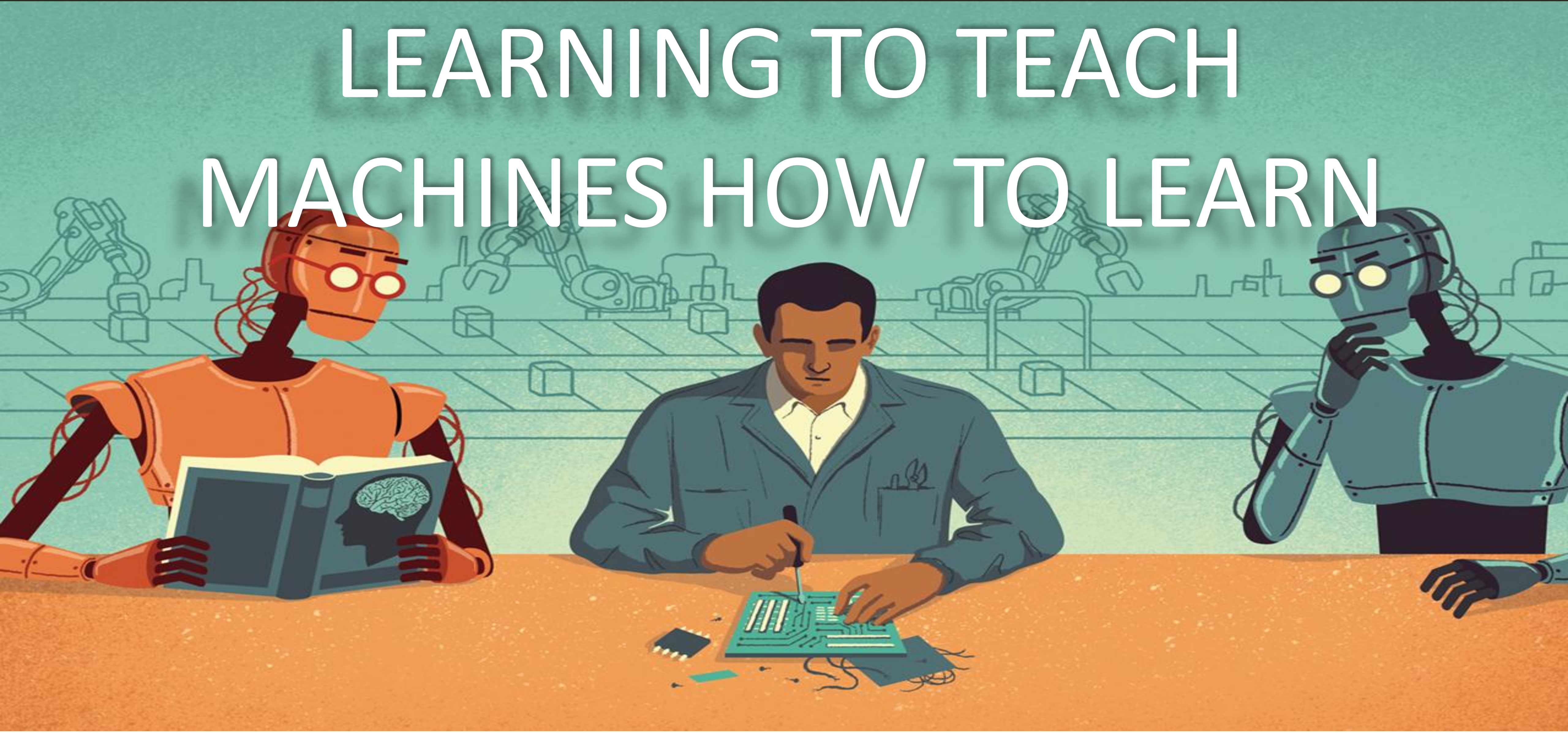
ASI

(Artificial Super Intelligence)

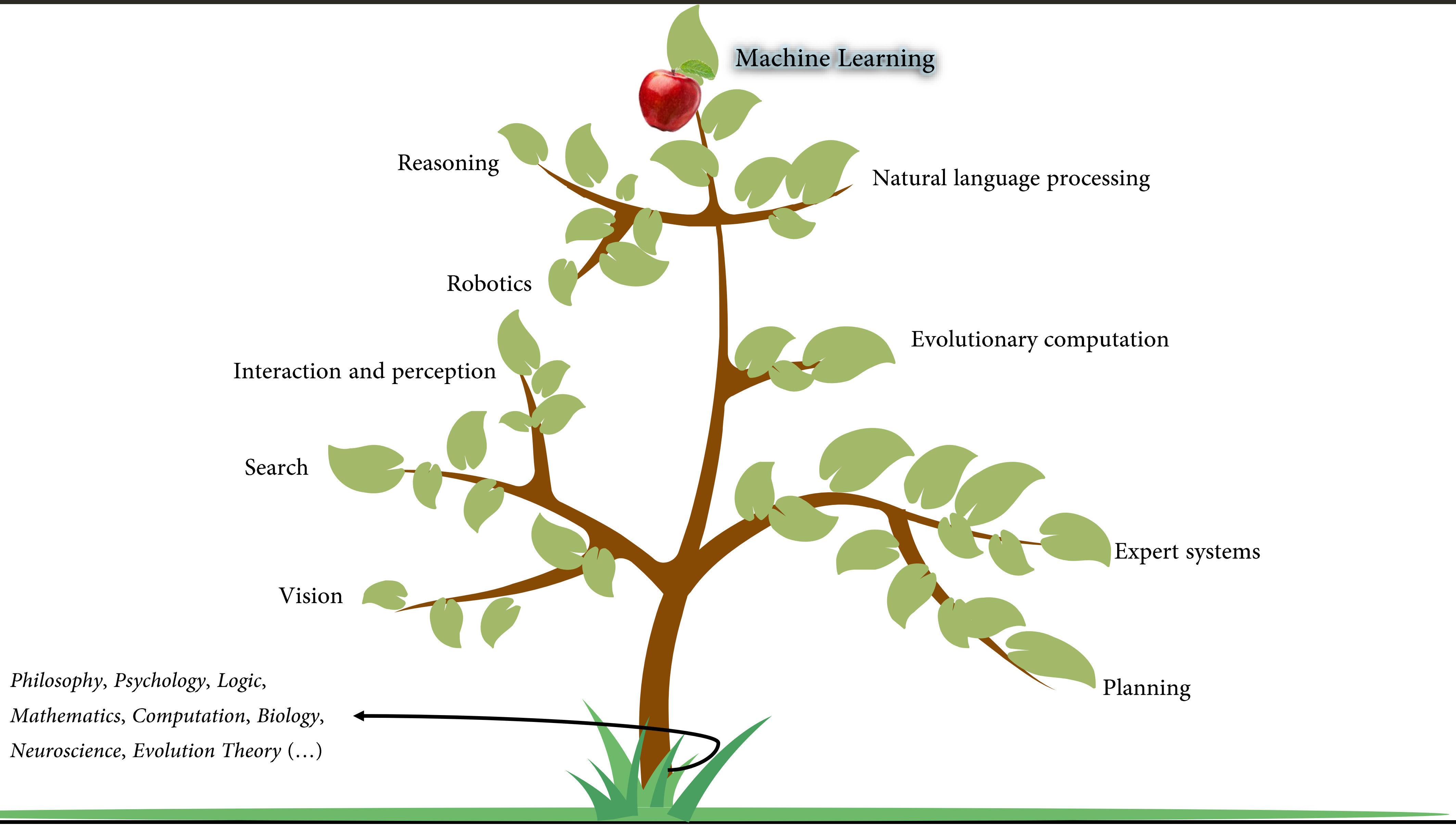
An **advanced AI that surpasses human intelligence and is fully self-aware**, with capabilities far beyond human cognitive limits.

Purely theoretical. Far off into the future, if at all ever possible.

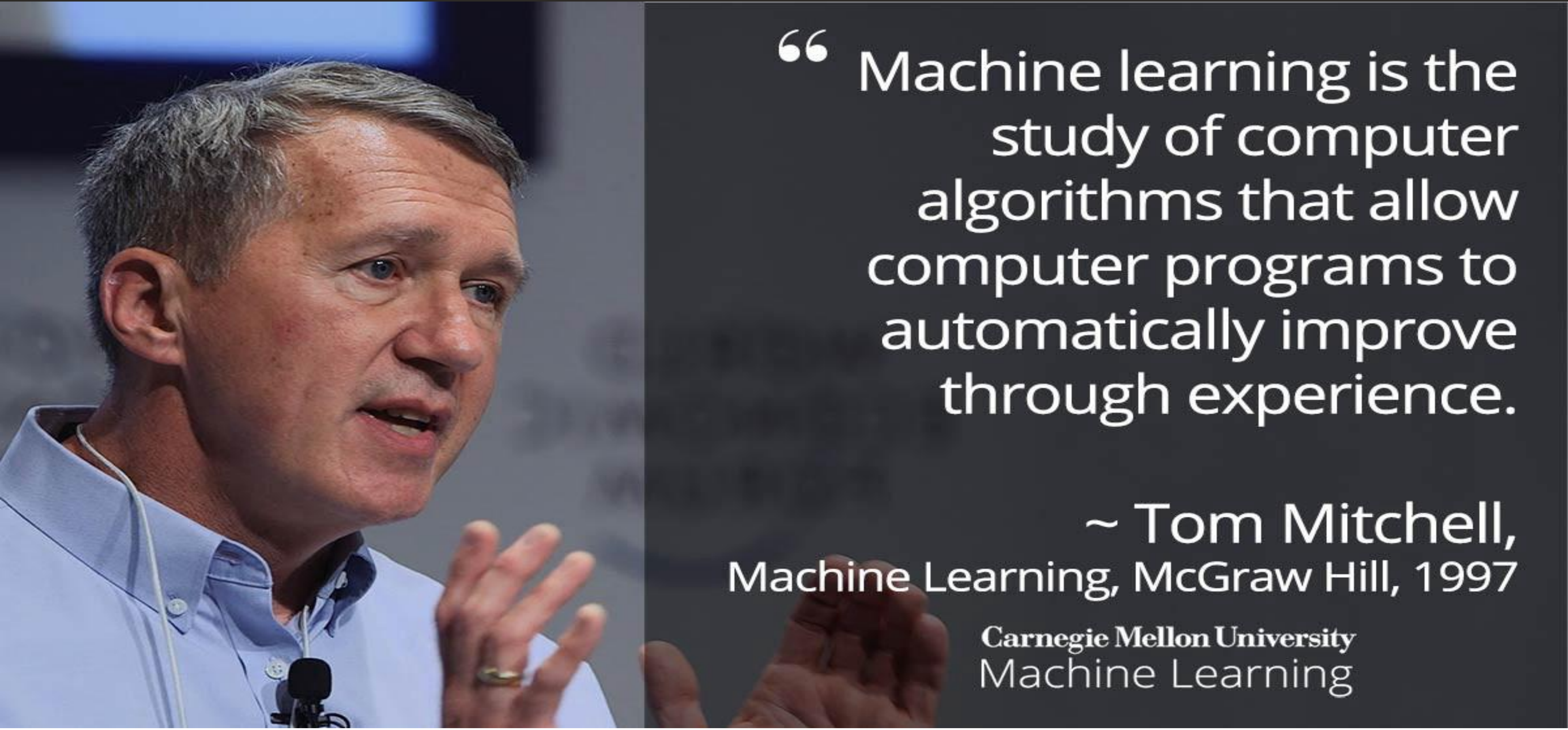
LEARNING TO TEACH MACHINES HOW TO LEARN



Machine Learning, the MVP of the AI field!



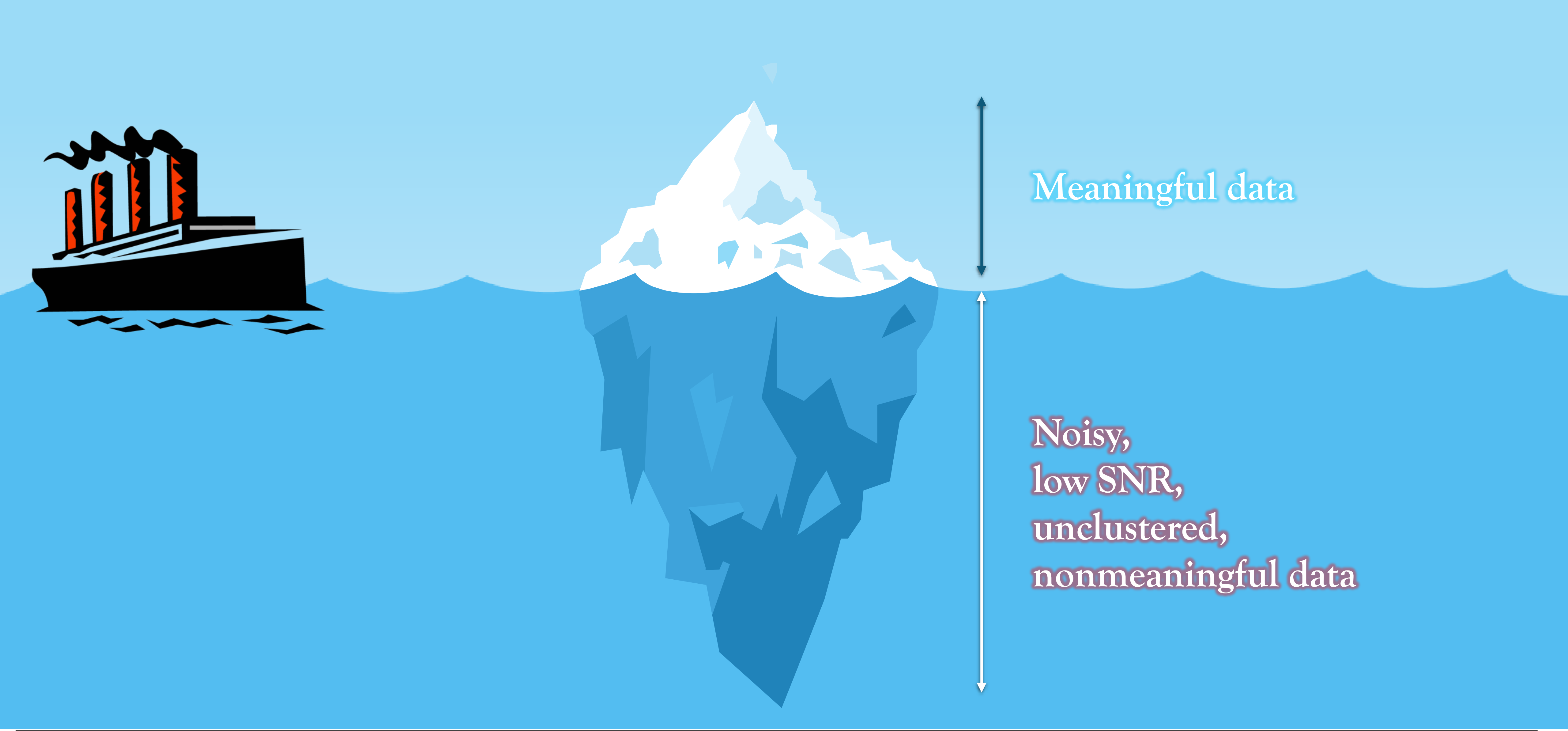
So, what is it then?



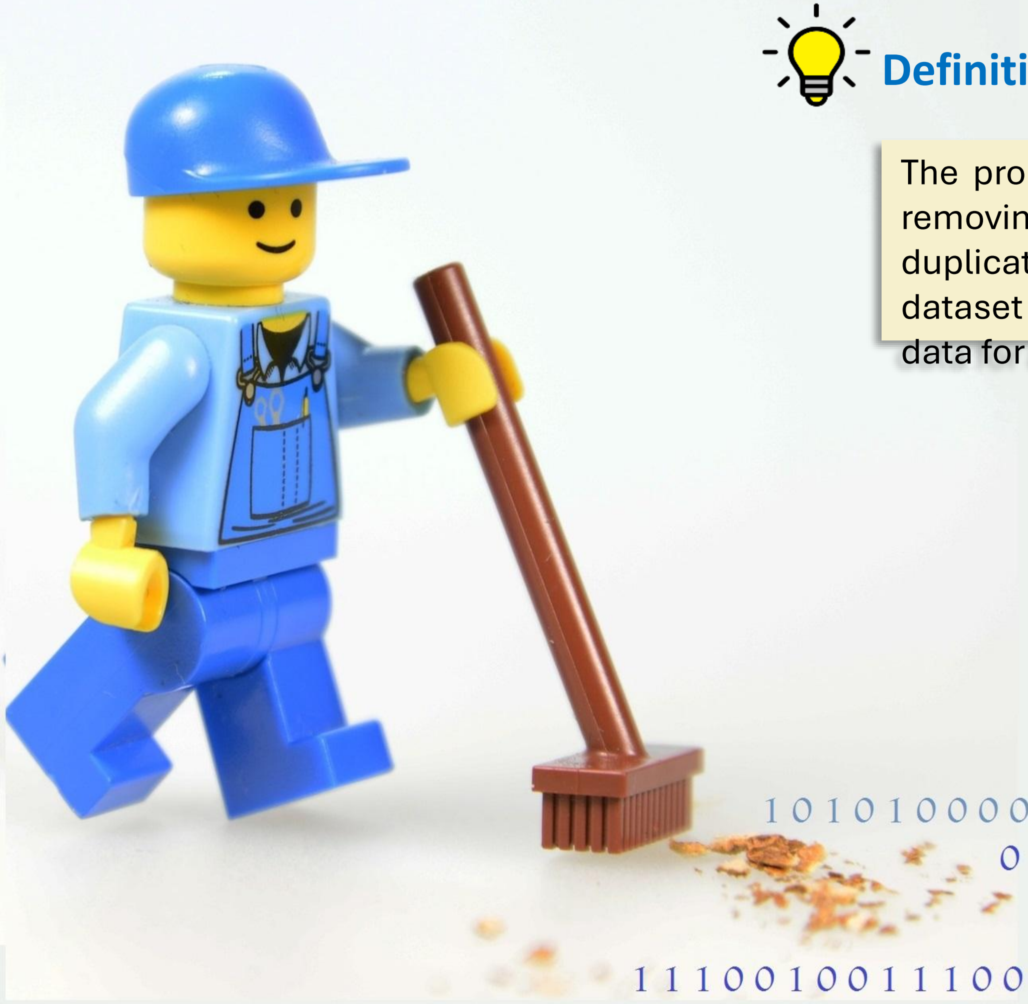
Definition

Extremely large and complex datasets that are too vast or varied to be managed and analyzed efficiently with traditional data processing tools, requiring advanced computational methods to extract meaningful insights.

Data = good results?



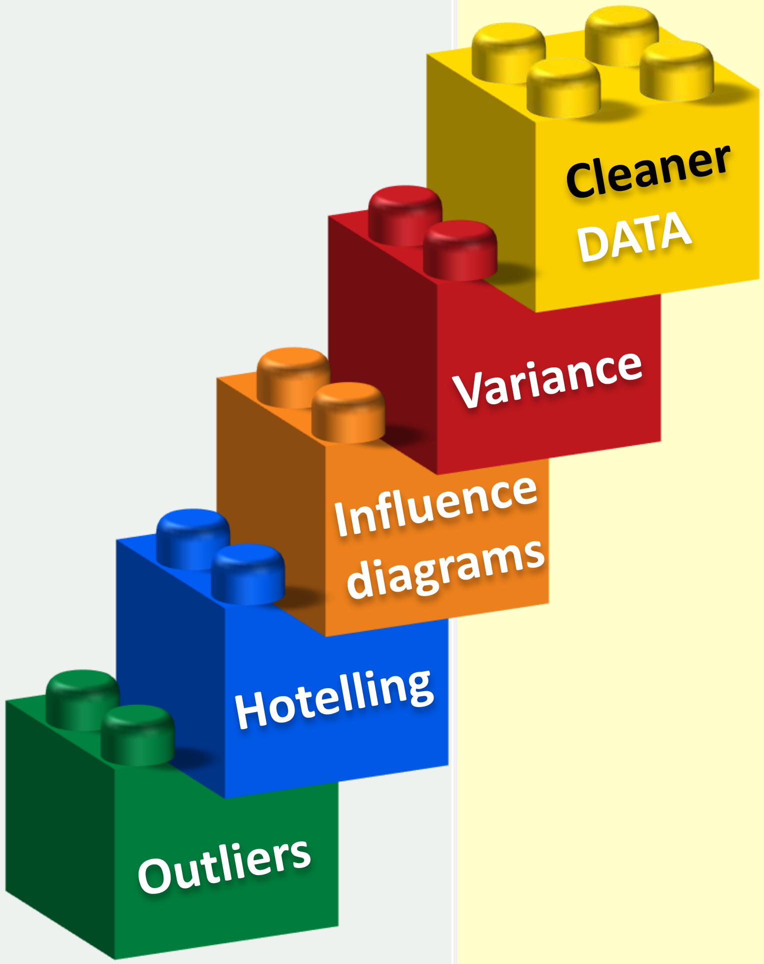
Nothing like a good Spring “data cleaning”



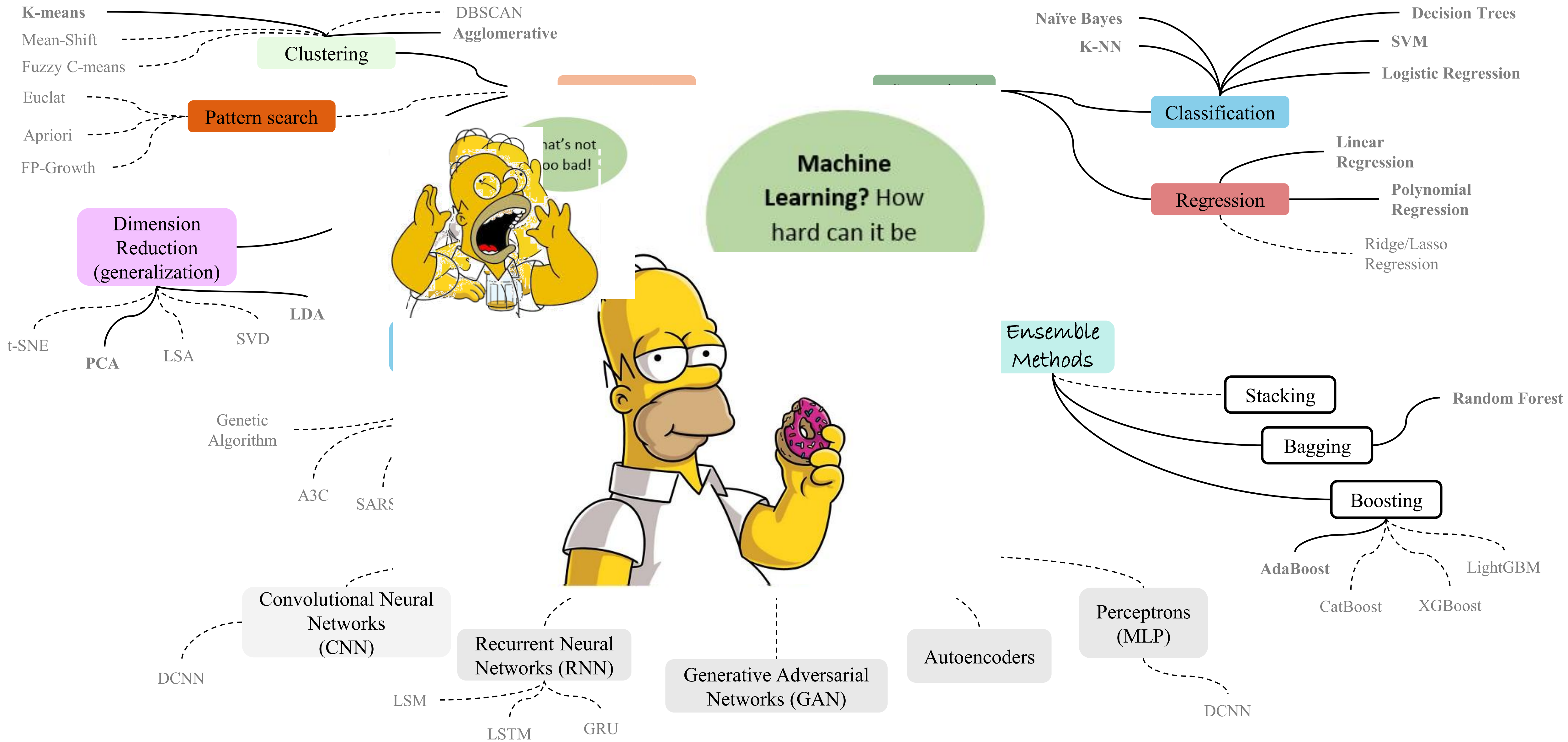
Definition

The process of identifying, correcting, or removing errors, inconsistencies, duplicates, or incomplete entries in a dataset to ensure high-quality, reliable data for analysis.

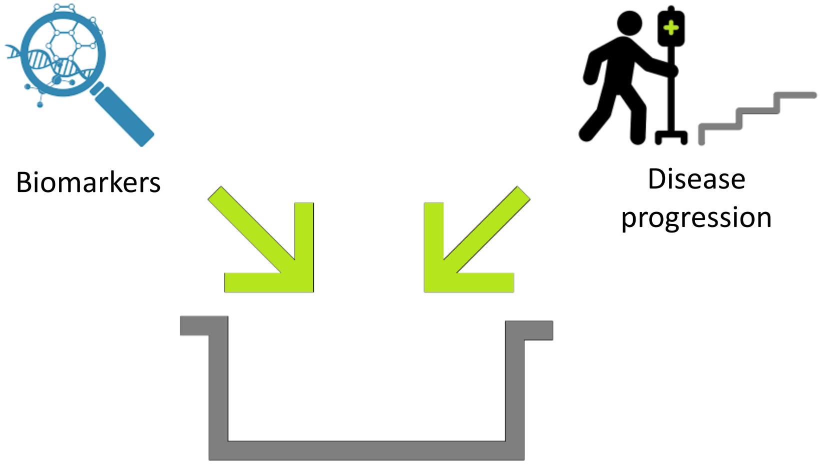
1 0 1 0 1 0 0 0 0 1 0 1 1 0 0 1
0 1 1 0 0 0 1 1 1 0 0 0 1 0
0 0 1 0 0 1 1 0 0 0 1 1 1 0 1 0
1 1 1 0 0 1 0 0 1 1 1 0 0 0 1 0 1
0 1 0 0 1 0 1 1 1 0 1 0 1 1 0 1
1 0 1 0 1 0 1 0 1 0 1 0



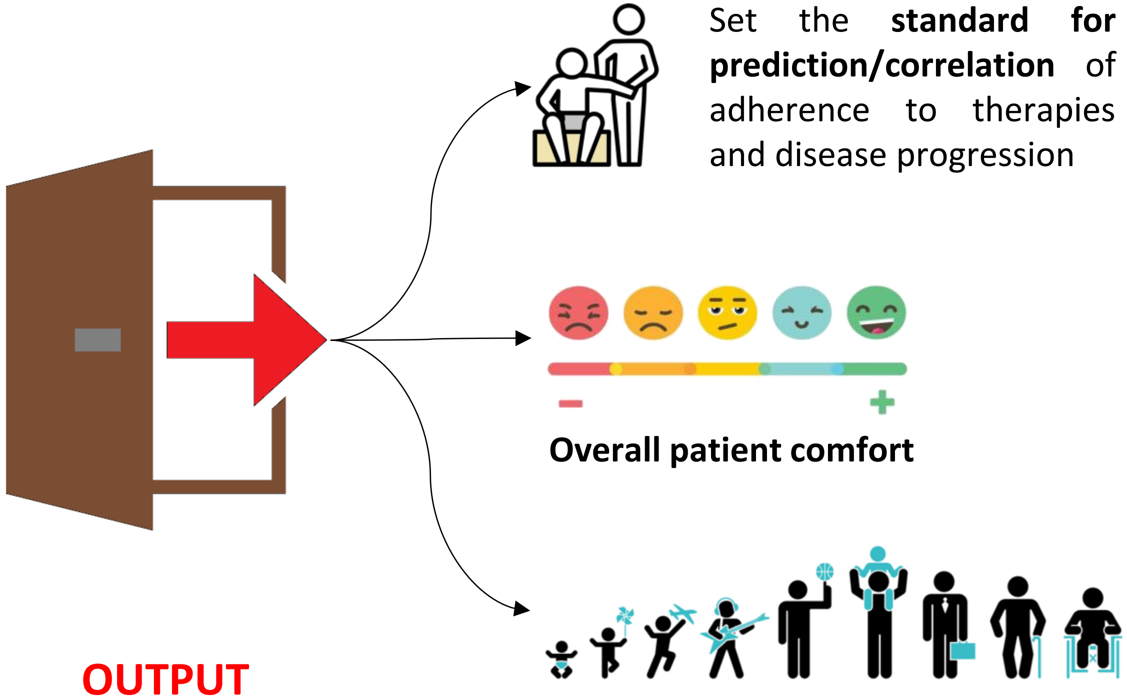
OK, now we know what ML and AI are. Easy, peasy! Right?...



The GOAL: clinical Impact – the patients' perspective

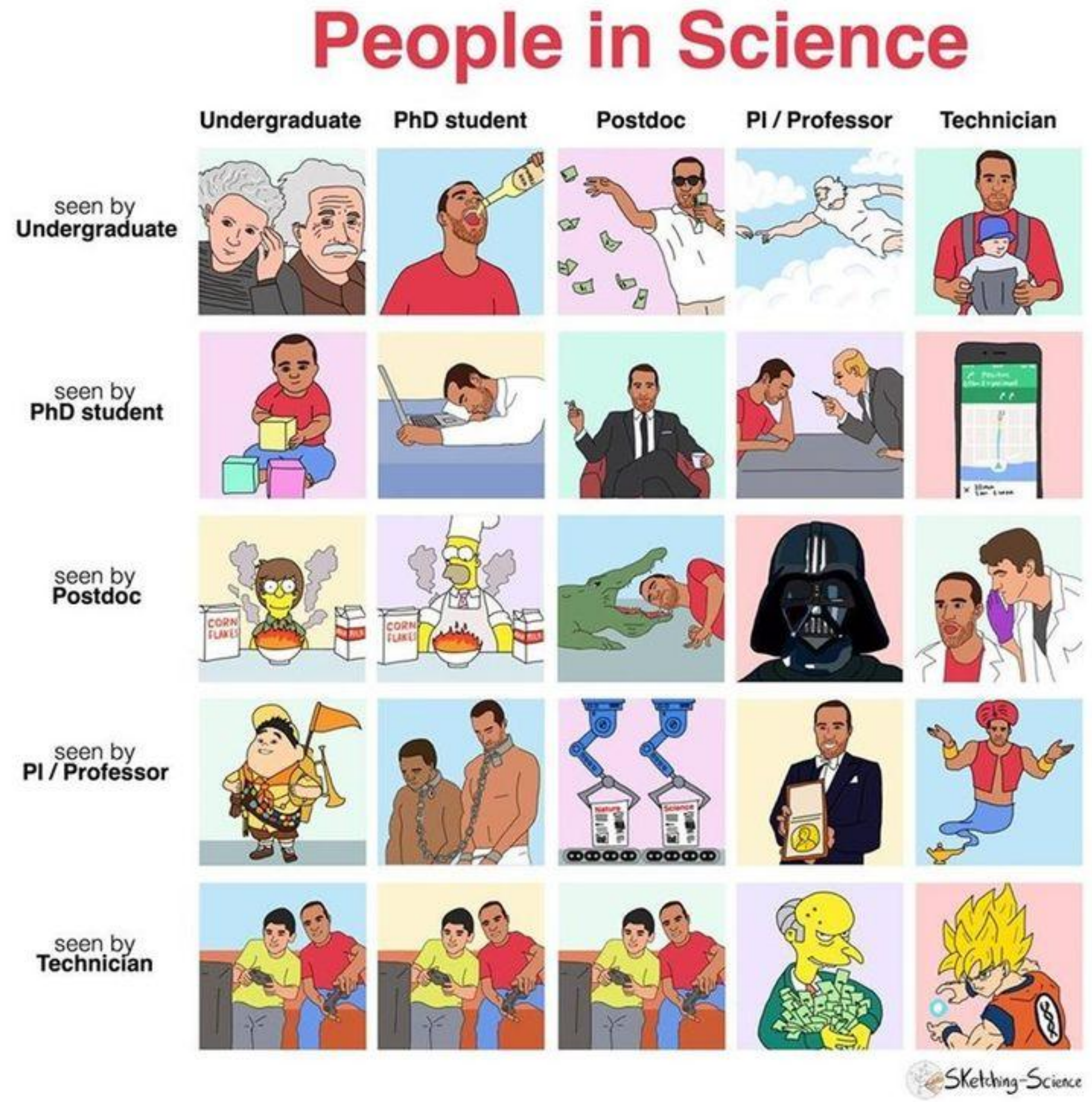


INPUT



OUTPUT

Set the **standard for prediction/correlation** of adherence to therapies and disease progression



Science!

If you want to check some of the work we do, you can check our Research Gate links below!

Rúben Araújo

Luís Ramalhete

ELSEVIER

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

Volume 255, 5 July 2021, 119680

A new method to predict genotoxic effects based on serum molecular profile

Rúben Araújo [a], Luís Ramalhete [a b], Hélder Paz [a], Carina Ladeira [c d e 1], Cecília R.C. Calado [a f 1]

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.saa.2021.119680

Get rights and content

Highlights

- Genotoxicity prediction based on serum molecular profile.
- Genotoxicity prediction by serum FTIR spectroscopic analysis.
- Biomonitoring based on a drop of blood analysis by FTIR spectroscopy.

The Whats & the Hows

Genotoxicity refers to the damage caused to genetic material by harmful agents, which can lead to mutations and cancer. Monitoring genotoxic effects in workers exposed to cytostatic drugs is critical for occupational safety. This study explored a fast, minimally invasive way to detect genotoxicity through FTIR spectroscopy of a single serum drop, offering a simpler alternative to conventional, labor-intensive assays.

Main Results

The method achieved 91% accuracy and strong specificity/sensitivity in distinguishing exposed from non-exposed individuals using a single serum drop.

ELSEVIER
Vibrational Spectroscopy
Volume 111, November 2020, 103177

Discriminating B and T-lymphocyte from its molecular profile acquired in a label-free and high-throughput method

Luís Ramalhete [a,b], Rúben Araújo [b], Cecília R.C. Calado [b,c]

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.vibspec.2020.103177 Get rights and content

Highlights

- B and T lymphocytes discrimination by a label-free method.
- A FTIR spectroscopy-based method to discriminate lymphocytes from peripheral blood.
- A SVM model based on second derivative spectra to predict B and T-lymphocytes.

The Whats & the Hows

B and T lymphocytes are key players in immune defense, and distinguishing them is essential in diagnosing infections, cancers, and autoimmune diseases. Traditional methods are costly and require labeling. This study applied FTIR spectroscopy to differentiate B and T cells based on their molecular fingerprints in a fast, label-free, and scalable manner.

Main Results

The approach reached 95% accuracy with a support vector machine model, revealing distinct molecular profiles between B and T cells.

NOVA MEDICAL SCHOOL

proteomes

IMPACT FACTOR 4.0 | Indexed in: PubMed | CITESCORE 6.5

Review

Proteomics for Biomarker Discovery for Diagnosis and Prognosis of Kidney Transplantation Rejection

Luís M. Ramalhete, Rúben Araújo, Aníbal Ferreira and Cecília R. C. Calado

MDPI

https://doi.org/10.3390/proteomes10030024

The Whats & the Hows

Organ rejection after kidney transplantation is a major cause of graft failure, driven by immune responses. Predicting rejection early can save lives and reduce treatment costs. This review highlights how <u>proteomics can identify biomarkers in urine or blood to detect rejection types</u> (acute, chronic, or subclinical) non-invasively and guide personalized immunotherapy.

Main Results

Emerging biomarkers show strong potential to detect early and specific immune rejection processes, enabling personalized treatment, though standardization is still needed.

methods and protocols

IMPACT FACTOR 2.3 | Indexed in: PubMed | CITESCORE 3.9

Article

Simplifying Data Analysis in Biomedical Research: An Automated, User-Friendly Tool

Rúben Araújo, Luís Ramalhete, Ana Viegas, Cristiana P. Von Rekowski, Tiago A. H. Fonseca, Cecília R. C. Calado and Luís Bento

MDPI

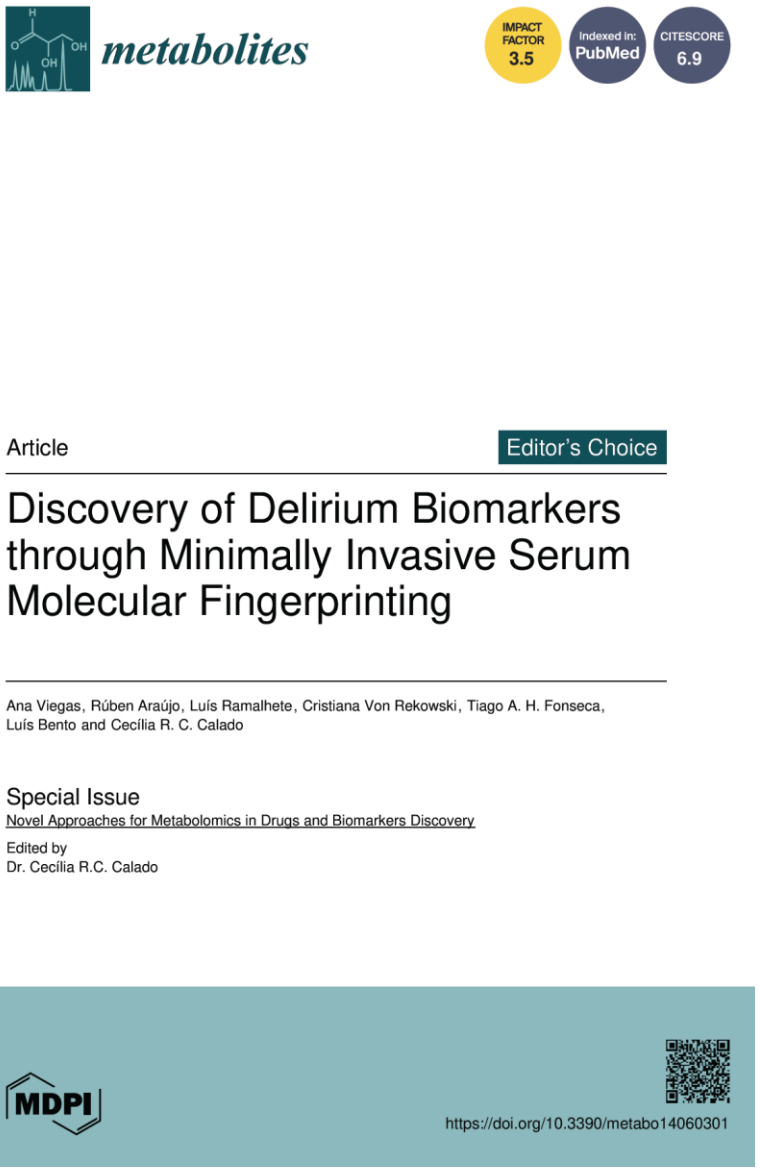
https://doi.org/10.3390/mps7030036

The Whats & the Hows

Many researchers struggle to clean and analyze complex biomedical datasets, slowing down discoveries. **ArsHive was developed as an open-source, user-friendly tool**—powered by GPT-4—to help non-experts preprocess, analyze, and report biomedical data efficiently. <u>Its aim is to democratize data science and accelerate research across disciplines</u>.

Main Results

It demonstrated effective normalization, unbiased reporting, and integrated AI assistance, easing complex clinical data processing for non-experts.



metabolites

IMPACT FACTOR 3.5 | Indexed in: PubMed | CITESCORE 6.9

Article | Editor's Choice

Discovery of Delirium Biomarkers through Minimally Invasive Serum Molecular Fingerprinting

Ana Viegas, Rúben Araújo, Luís Ramalhete, Cristiana Von Rekowski, Tiago A. H. Fonseca, Luís Bento and Cecília R. C. Calado

Special Issue
Novel Approaches for Metabolomics in Drugs and Biomarkers Discovery

Edited by
Dr. Cecília R.C. Calado

MDPI

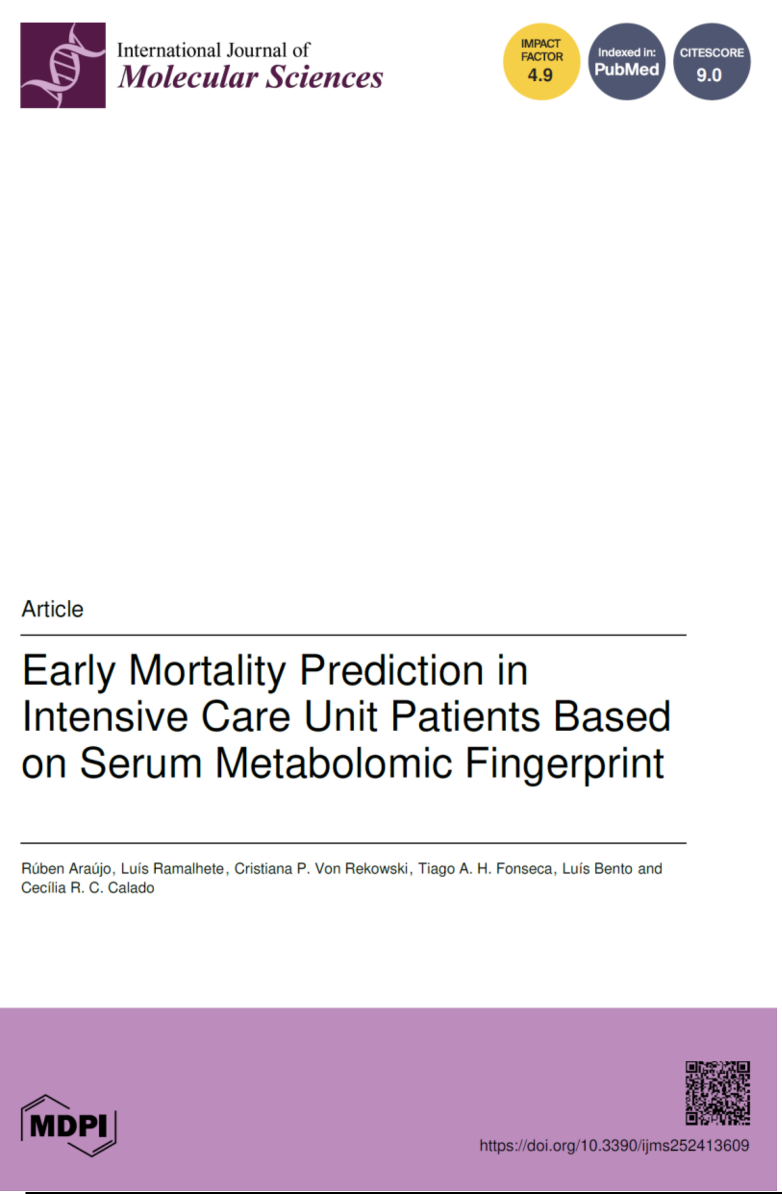
https://doi.org/10.3390/metabo14060301

The Whats & the Hows

Delirium is a sudden and severe disturbance in mental function, often seen in ICU patients, especially during COVID-19. It is underdiagnosed and frequently mistaken for other conditions. This study used FTIR spectroscopy of serum to identify molecular fingerprints that could <u>predict delirium</u>, providing a rapid, minimally invasive diagnostic tool to improve patient outcomes.

Main Results

Using selected spectral bands, the model achieved AUC, sensitivity, and specificity above 0.92, offering a fast and scalable delirium biomarker strategy.



International Journal of *Molecular Sciences*

IMPACT FACTOR 4.9 | Indexed in: PubMed | CITESCORE 9.0

Article

Early Mortality Prediction in Intensive Care Unit Patients Based on Serum Metabolomic Fingerprint

Rúben Araújo, Luís Ramalhete, Cristiana P. Von Rekowski, Tiago A. H. Fonseca, Luís Bento and Cecília R. C. Calado

MDPI

https://doi.org/10.3390/ijms252413609

The Whats & the Hows

Predicting mortality risk in ICU patients is crucial for early intervention and optimal care. During pandemics or resource-limited situations, accurate prediction tools can save lives. This study used FTIR-based serum analysis to model metabolic changes over time, supporting early and scalable mortality risk prediction.

Main Results

High prediction performance (AUC up to 0.98) was achieved using spectral bands, supporting real-time monitoring with standard blood samples.

metabolites

IMPACT FACTOR 3.7 | Indexed in: PubMed | CITESCORE 6.9

Article

Cytokine-Based Insights into Bloodstream Infections and Bacterial Gram Typing in ICU COVID-19 Patients

Rúben Araújo, Luís Ramalhete, Cristiana P. Von Rekowski, Tiago A. H. Fonseca, Cecília R. C. Calado and Luís Bento

Special Issue
Towards Clinical Interpretation of Metabolomic Data

Edited by
Prof. Dr. Tomáš Adam

MDPI

https://doi.org/10.3390/metabo15030204

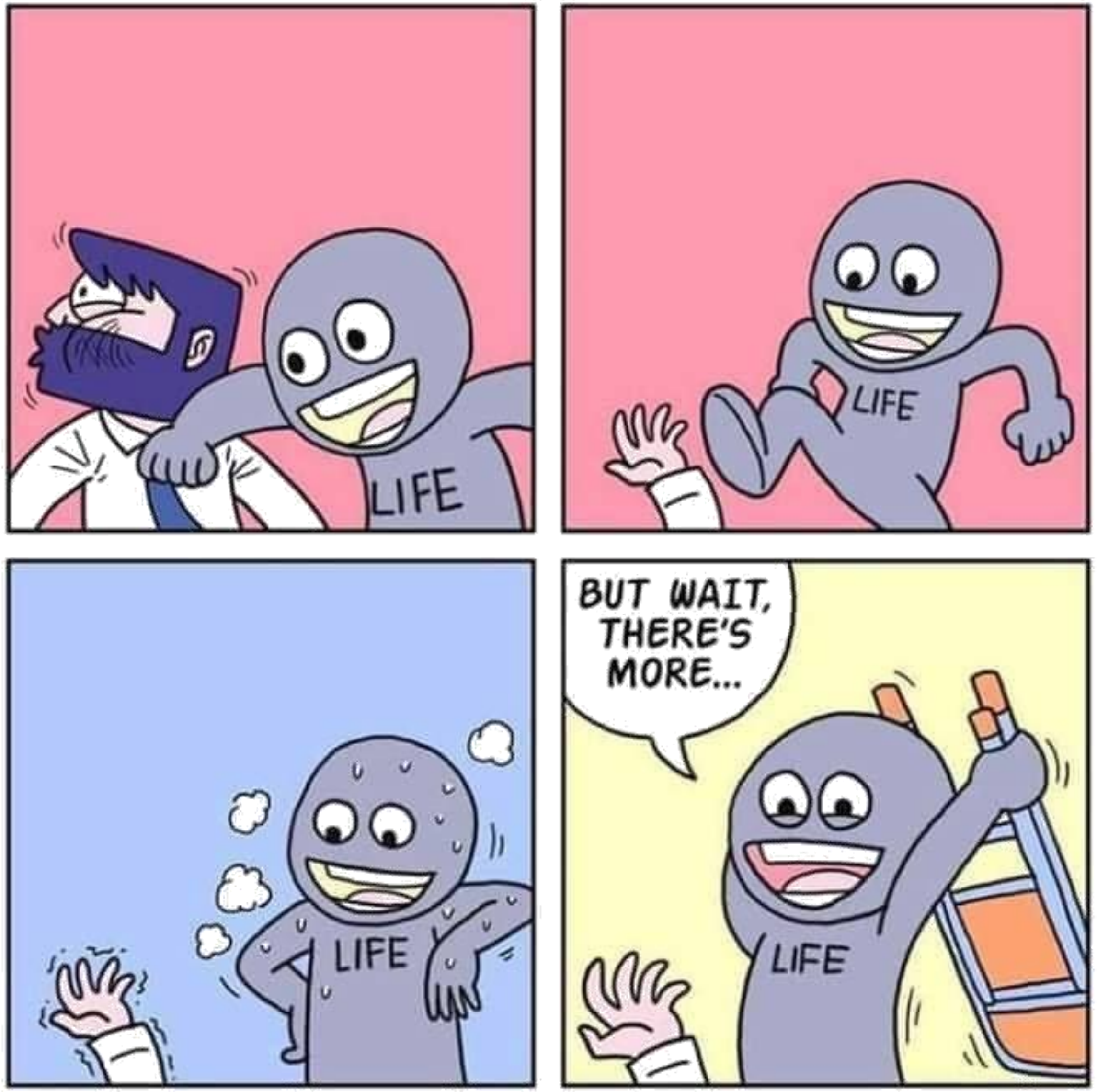
The Whats & the Hows

Bloodstream infections (BSIs) are life-threatening, especially in ICU patients with weakened immunity. Rapid identification is vital to guide treatment and prevent death. This study used cytokine profiling and machine learning to detect infections and identify bacterial Gram type from serum samples, aiming to improve early diagnosis and antibiotic precision.

Main Results

Models reached AUCs up to 0.98, showing cytokine-based diagnostics can significantly improve infection management in critical care.

@BANGGARANG.COMICS

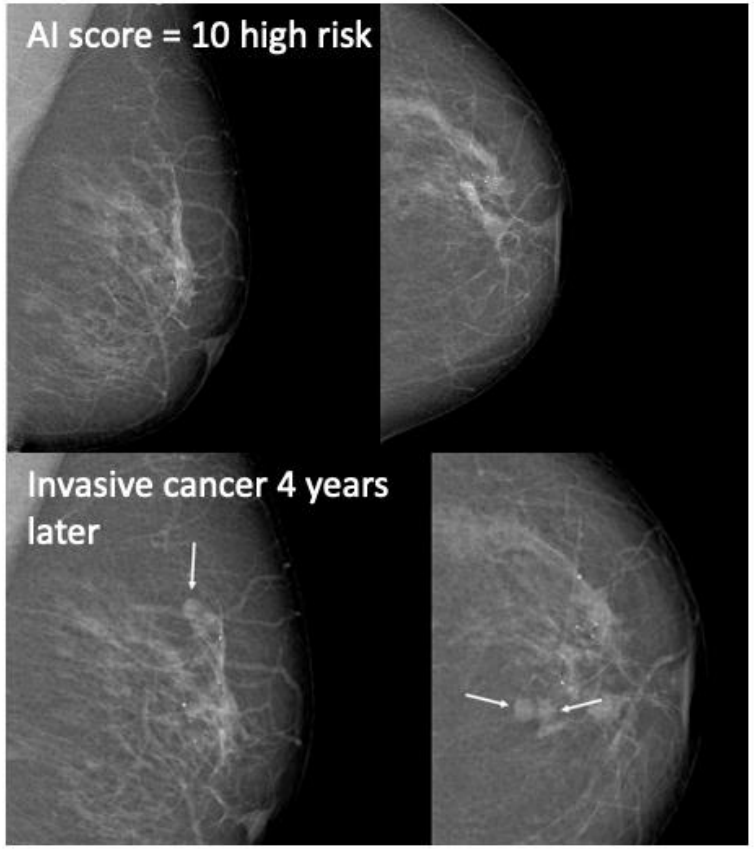


But wait! There's more!

Other examples of BIG DATA in Healthcare....

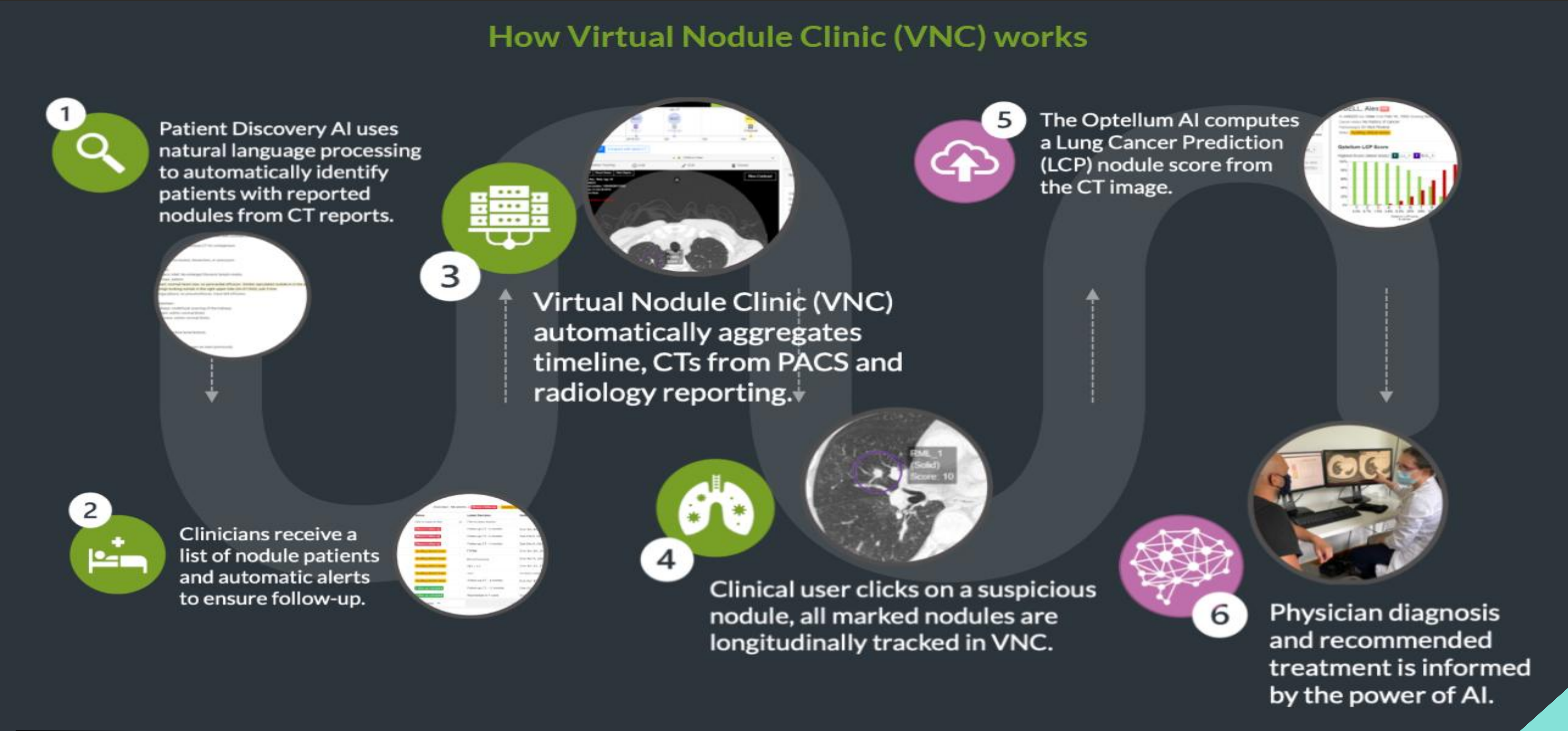
MEDICAL IMAGING

AI Score on Screening Mammograms Preceding Breast Cancer Diagnosis

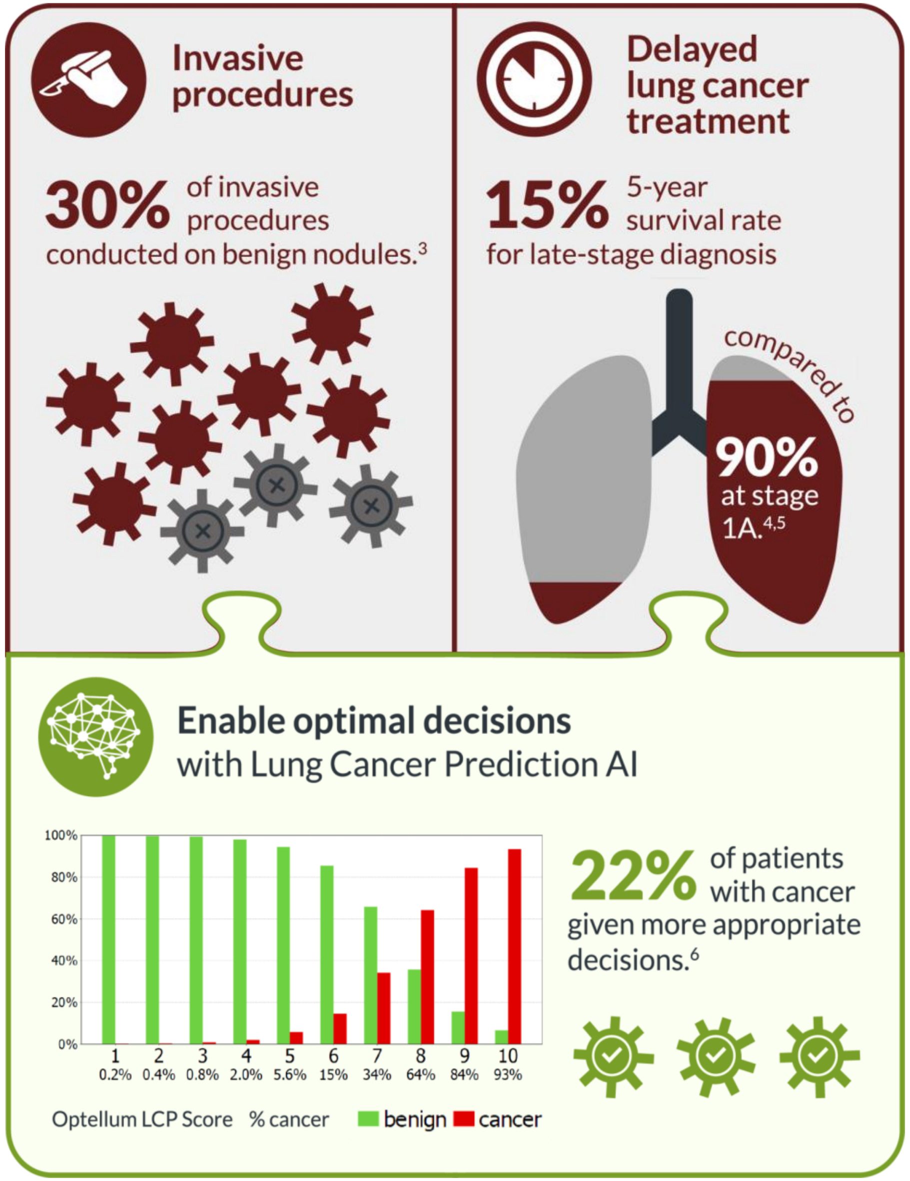


- Retrospective study of 1602 patients with screen-detected or interval breast cancers; AI model assessed risk on preceding negative mammograms from 1 to 10, with 10 being high risk.
- 38% (389 of 1016) of screen-detected and 39% (231 of 586) of interval cancers had the highest AI risk score on the prior screening mammogram.
- High-risk features like density with calcifications were more frequent on high-risk mammograms than low risk, 14% (43 of 317) and 5% (15 of 322), respectively, in cases of subsequent invasive screen-detected cancers.

Larsen M et al. Published Online: October 17, 2023
https://doi.org/10.1148/radiol.230989
Radiology

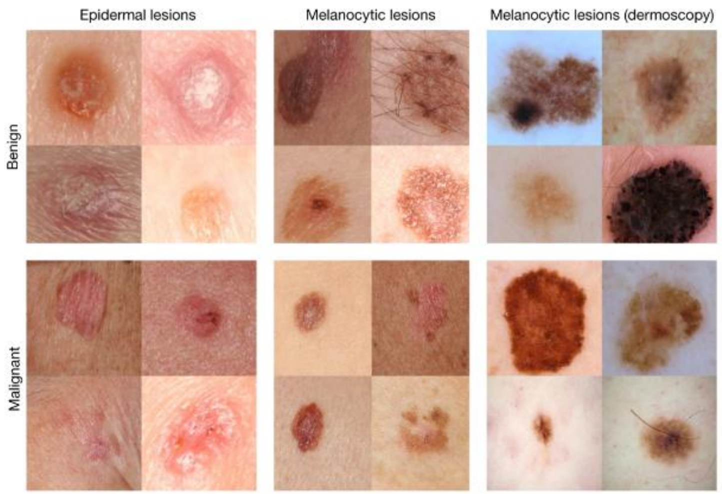

How Virtual Nodule Clinic (VNC) works
1
Patient Discovery AI uses natural language processing to automatically identify patients with reported nodules from CT reports.
2
Clinicians receive a list of nodule patients and automatic alerts to ensure follow-up.
3
Virtual Nodule Clinic (VNC) automatically aggregates timeline, CTs from PACS and radiology reporting.
4
Clinical user clicks on a suspicious nodule, all marked nodules are longitudinally tracked in VNC.
5
The Optellum AI computes a Lung Cancer Prediction (LCP) nodule score from the CT image.
6
Physician diagnosis and recommended treatment is informed by the power of AI.


Optellum
Optellum Receives FDA Clearance for the World's First AI-Powered Clinical Decision Support Software for Early Lung Cancer Diagnosis
FDA CLEARED
@optellum
www.optellum.com


Invasive procedures
30% of invasive procedures conducted on benign nodules.3
Delayed lung cancer treatment
15% 5-year survival rate for late-stage diagnosis
compared to
90% at stage 1A.4,5
Enable optimal decisions with Lung Cancer Prediction AI
100%
80%
60%
40%
20%
0%
1 0.2%
2 0.4%
3 0.8%
4 2.0%
5 5.6%
6 15%
7 34%
8 64%
9 84%
10 93%
Optellum LCP Score
% cancer
benign
cancer
22% of patients with cancer given more appropriate decisions.6

DIAGNOSTIC Apps

AI in dermatology



2013

20 % of Classification accuracy (CA)

2017

50 % CA

Nature (2017)

2018

60 % CA

Annals of Oncology (2018)

2019

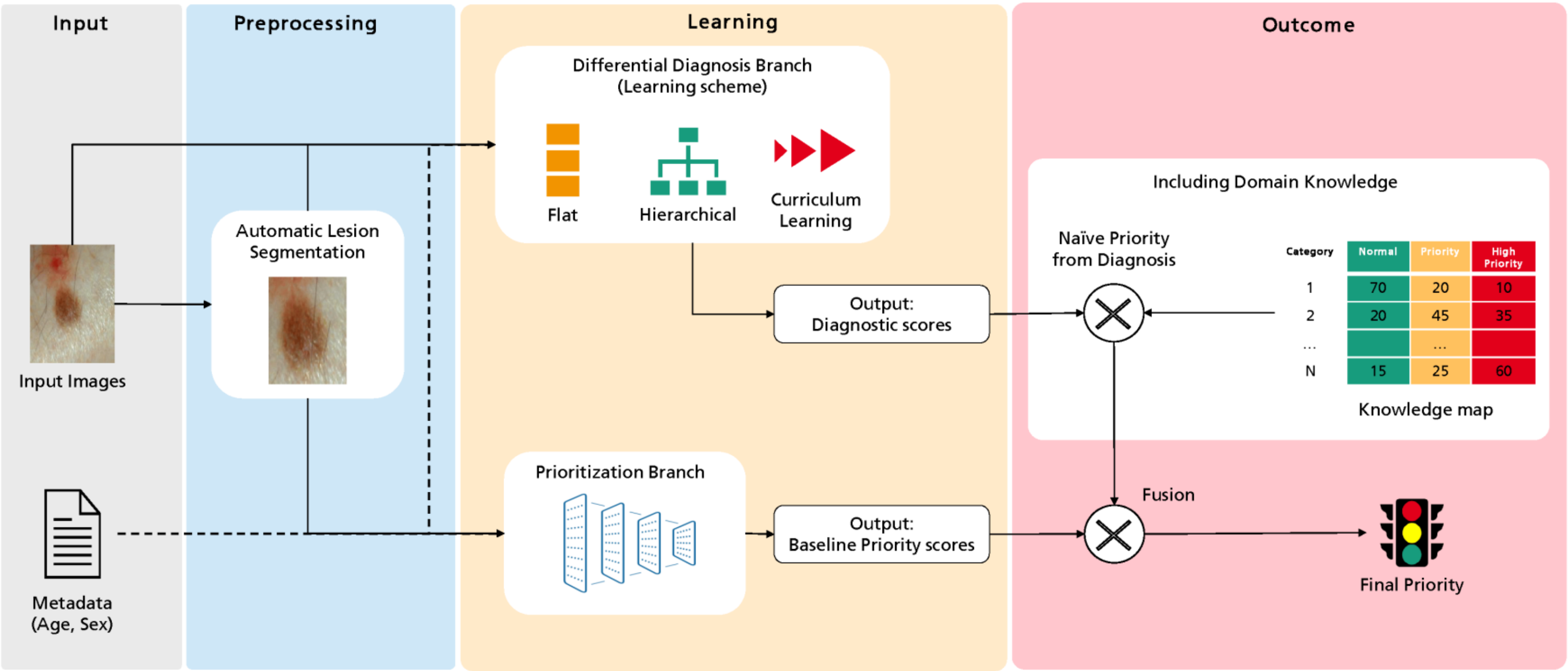
95% CA

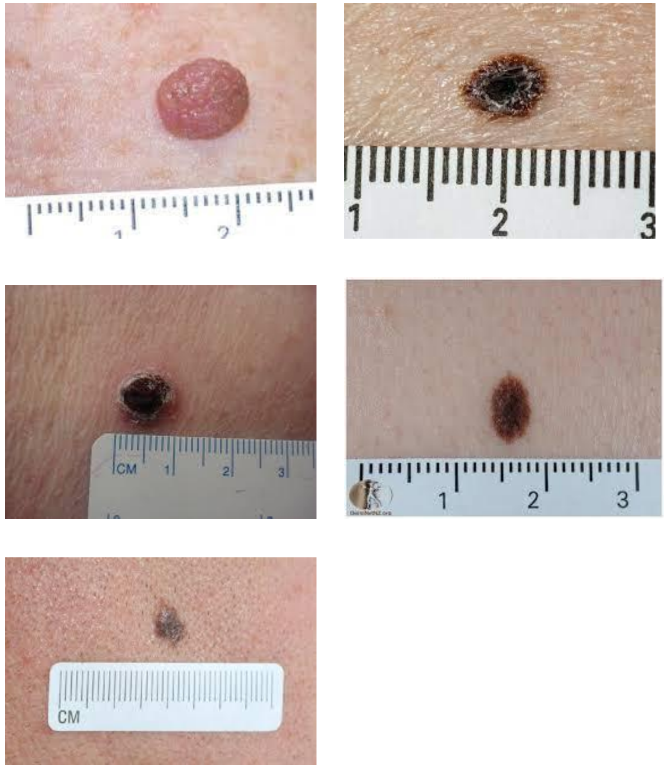
JAMA Dermat (2019)

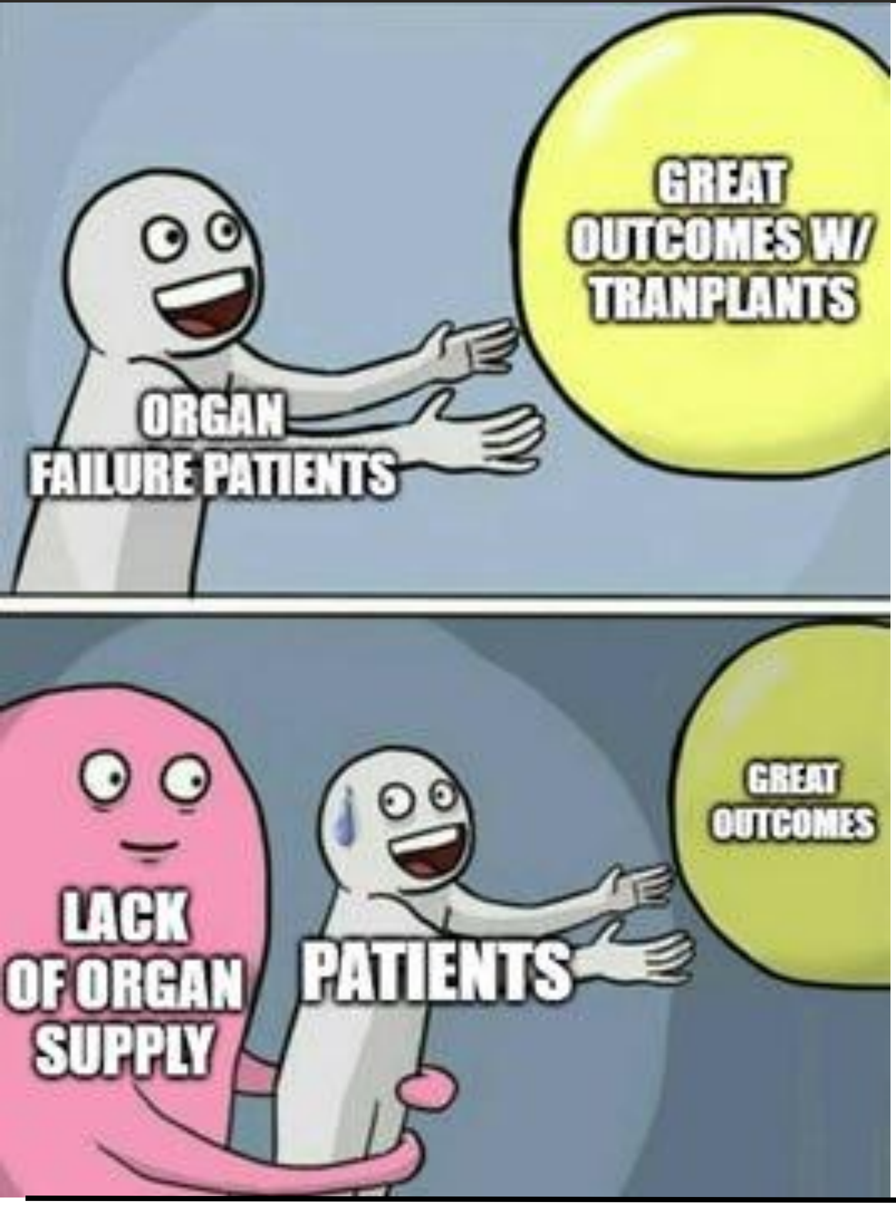
2022

95 % CA

Clinical Trial


Input
Preprocessing
Learning
Outcome
Differential Diagnosis Branch
(Learning scheme)
Flat
Hierarchical
Curriculum Learning
Automatic Lesion Segmentation
Input Images
Metadata (Age, Sex)
Output: Diagnostic scores
Prioritization Branch
Output: Baseline Priority scores
Including Domain Knowledge
Naïve Priority from Diagnosis
Category
Normal
Priority
High Priority
1
70
20
10
2
20
45
35
...
...
N
15
25
60
Knowledge map
Fusion
Final Priority





07 Organ Transplantation

Donation must be strictly voluntary

Donation is always voluntary, and must respect the donor's freedom of choice.

No payment allowed for organs

Organ donation cannot be monetized or used for commercial purposes — it's strictly altruistic.

Opt-out system: default donor status for residents

All Portuguese nationals, stateless persons, and foreigners living in Portugal are considered potential organ donors after death, unless they've officially opted out.

Must register in RENNDA[2] to refuse donation

To not donate, individuals (or legal guardians for minors/incapacitated persons) must register themselves in the RENNDA (*Registo Nacional de Não Dadores*).

Must respect human dignitiy and rights throughout the process

Donations must preserve respect for the donor's and recipient's human dignity and fundamental rights, adhering to ethical EU and WHO standards.

[1] https://diariodarepublica.pt/dr/detalhe/lei/12-1993-692651; [2] https://www.ipst.pt/index.php/pt/rennda

A Complex Law Timeline: an Overly Simplified Version [1]

1993

Lei nº. 12/93: Establishes the first legal framework for organ and tissue transplantation in Portugal. Introduces presumed consent for organ donation.

1994-2000

Establishment of the National Register of Non-Donors and the CEDACE (National Center for Bone Marrow and Stem Cell Donors). Approval of death verification criteria and donor-recipient matching procedures.

2001-2009

Ratification of key European bioethics and human rights conventions. Adoption of EU directives to standardize quality, safety, and financing mechanisms for transplantation.

Lei nº. 12/2009: Updates safety standards for tissues and cells.

2010-2015

Creation of the National Renal Cross Donation Program.

Lei nº. 36/2009: and laws 1/2015 & 2/2015: Reinforce EU-aligned quality and safety regimes for human organ transplantation.

Decreto Lei nº. 168/2015: Establishes protections for living donors.

2017-2020

Lei nº. 99/2017: Amends previous tissue and cell safety laws.

Portugal ratifies international conventions on trafficking in human organs.

Despacho nº. 8262/2020: 20 July established as the National Day for Organ Donation and Transplantation.

2025

Despacho nº. 3391/2025: Launches an Action Plan under the National Strategy for Renal Health Promotion and Chronic Kidney Disease Care (2023–2026).

[1] https://www.ipst.pt/index.php/pt/legis-transplanatcao

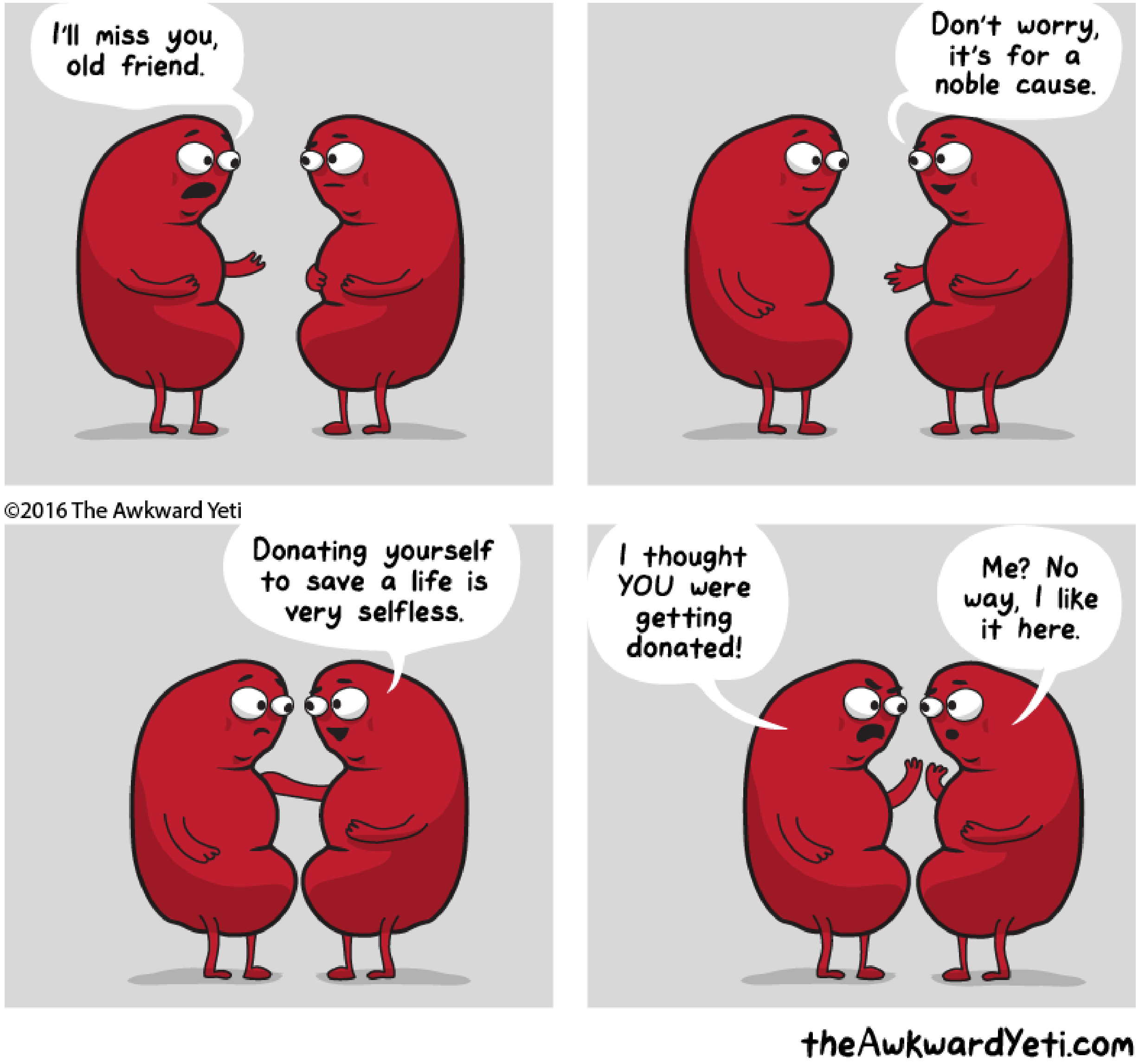
AI Strategies to cope with organ shortage

Table 1. Overview of ML and AI applications to mitigate organ shortage.

Problem	Organ Type	Population	ML AI Models	Results	Ref.
Donor identification	Donor	Potential organ donors (n = 80) Not potential organ donors (n = 564)	Neural networks	AUC-ROC 0.97, sensitivity 0.84, specificity 0.93	[54]
Optimizing the consent rate	Donor	Consent (n = 1461) No consent (n = 2811)	Networked Logistic Regression	Accuracy 99.912, precision 0.999, recall 0.999, F-Measure 0.999	[52]
Donor organ quality	Kidney	Kidney transplants (n ≈ 60,000)	Random forest	VIMP 0.0087	[56]
Donor discard	Liver	Organ used (n = 167,676) Organ discarded (n = 56,422)	XGBoost	AUC-ROC 0.93, AUC-PR 0.87, and F1 statistic 0.76	[57]
	Kidney	Organ used (n = 184,746) Organ discarded (n = 41,965)		AUC-ROC 0.95, AUC-PR 0.88, and F1 statistic of 0.79	
Kidney discard	Kidney	Organ used (n = 61,313) Organ discarded (n = 12,510)	Random forest	AUC-ROC 0.90 and balanced accuracy 0.78	[58]
Kidney discard	Kidney	Organ used (n = 79,039) Organ discarded (n = 23,207)	Logistic regression	C statistic 0.89	[59]
Optimizing organ yield	Donor	Donors (n = 89,520)	Tree-based gradient boosting	MAE 0.73, MSE 0.87	[60]
Decision to accept	Kidney	Accepted kidney transplants (n = 36,653)	Neural networks	AUC-ROC 0.81, F1-score 0.66	[61]

AUC-ROC, Receiver operating characteristic curve; VIMP, permutation variable importance; AUC-PR, Precision-Recall Area Under the Curve; MAE, Mean Average Error; MSE, Mean Squared Error.

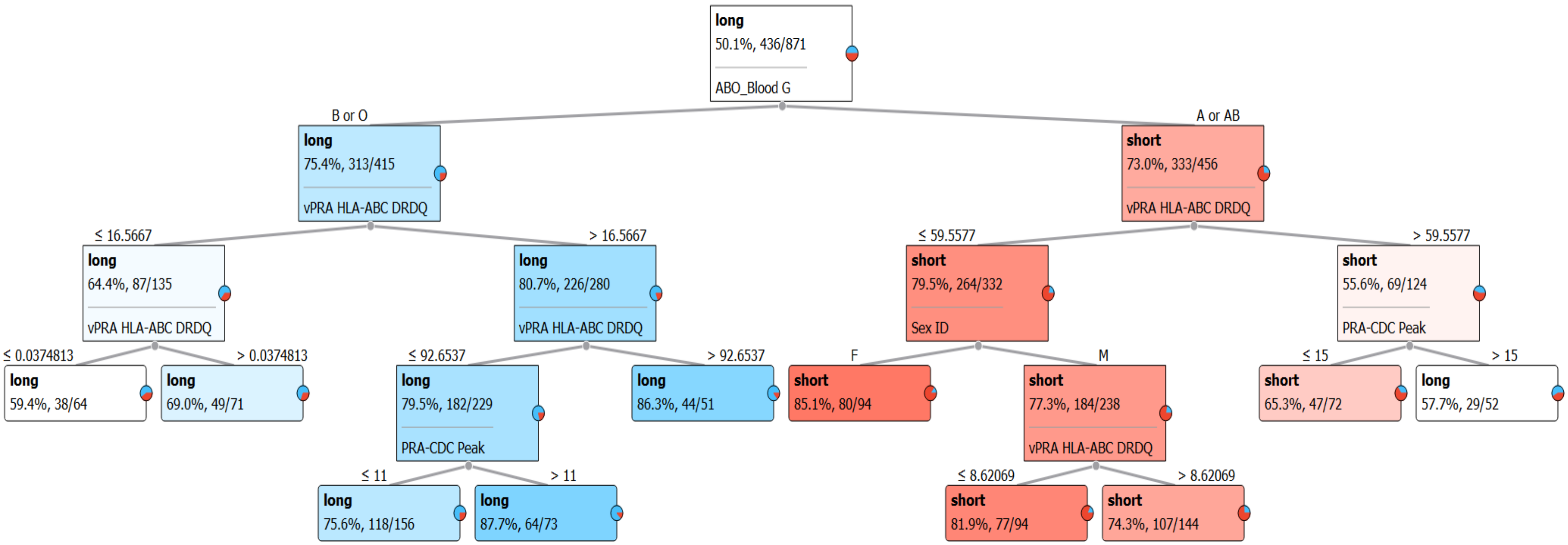
Ramalhete *et al*, (2024) Revolutionizing Kidney Transplantation: Connecting Machine Learning and Artificial Intelligence with Next-Generation Healthcare—From Algorithms to Allografts



Transplantation

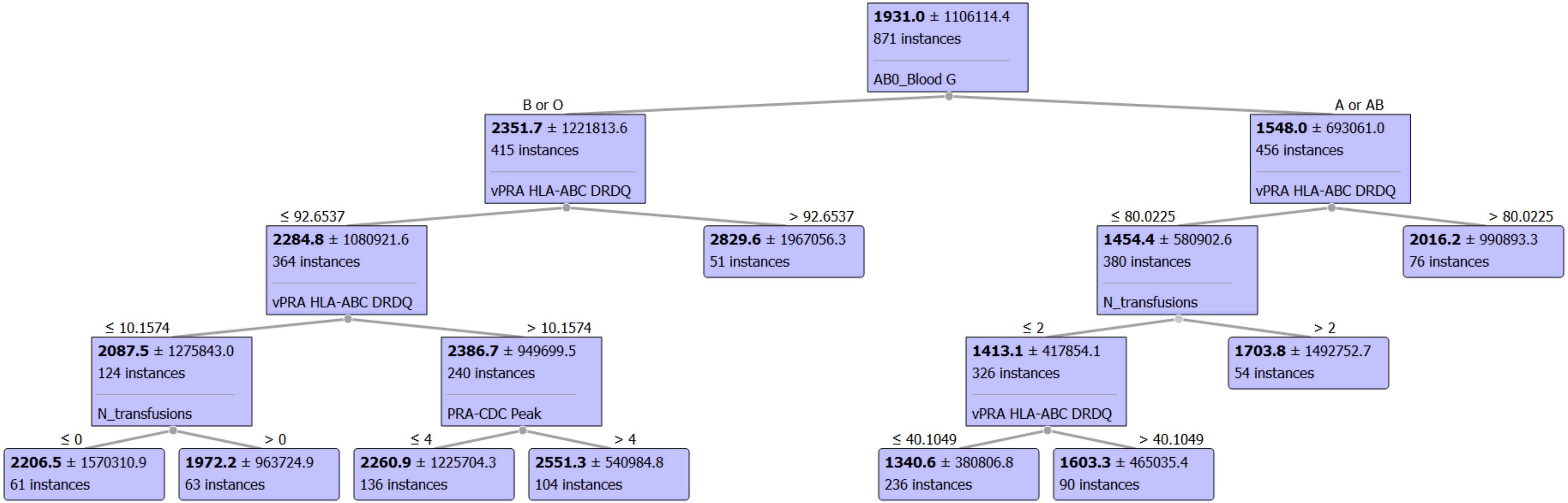
Who is transplanted in the waiting list?

Based on median waiting time

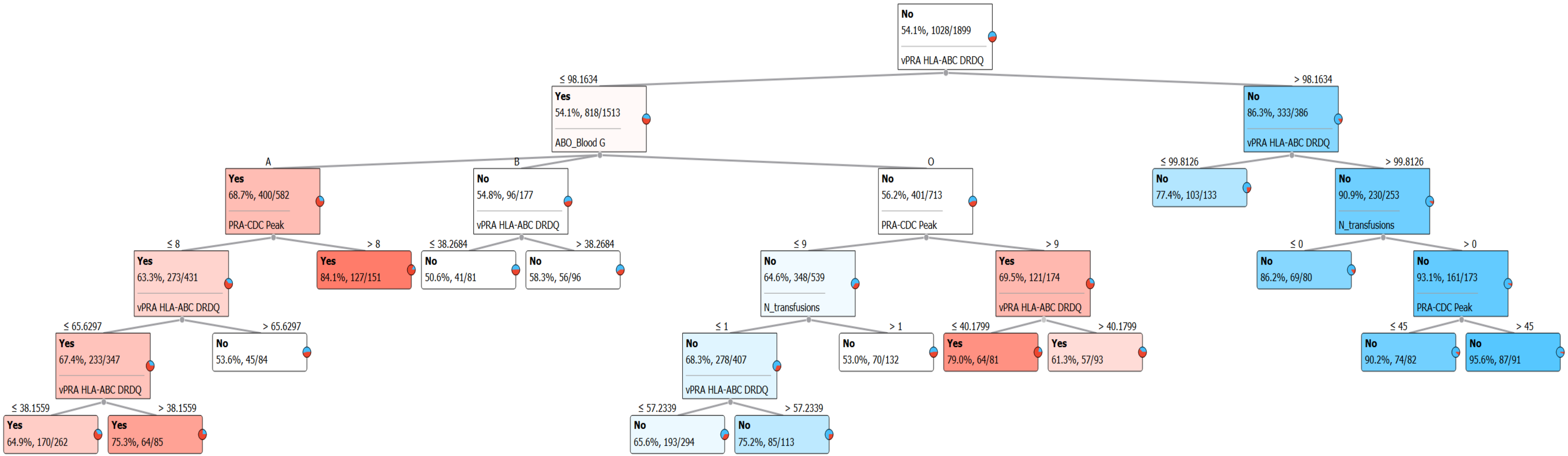


Who is transplanted in the waiting list?

Based on average waiting time

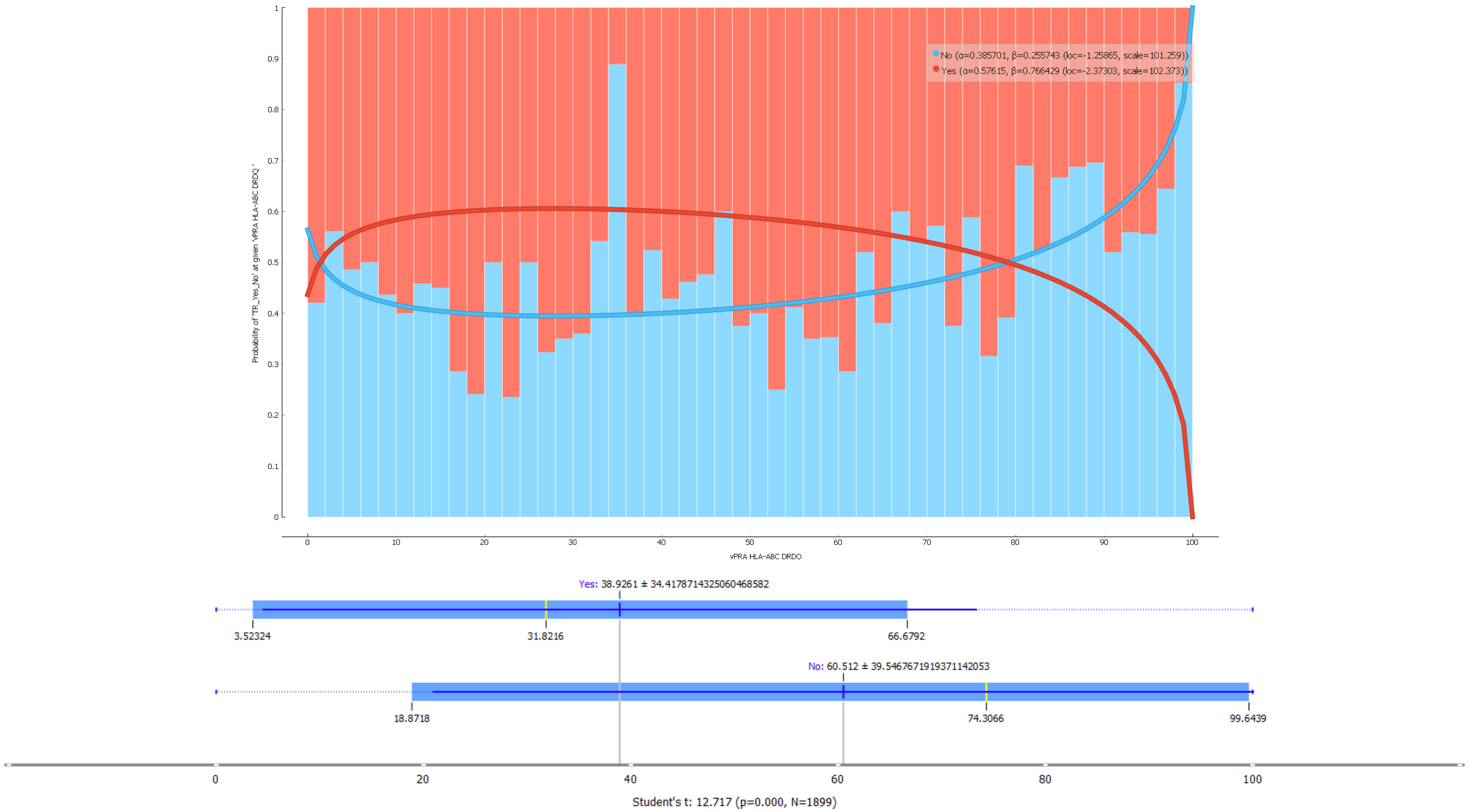


Who stays in the waiting list *vs* transplanted patients?



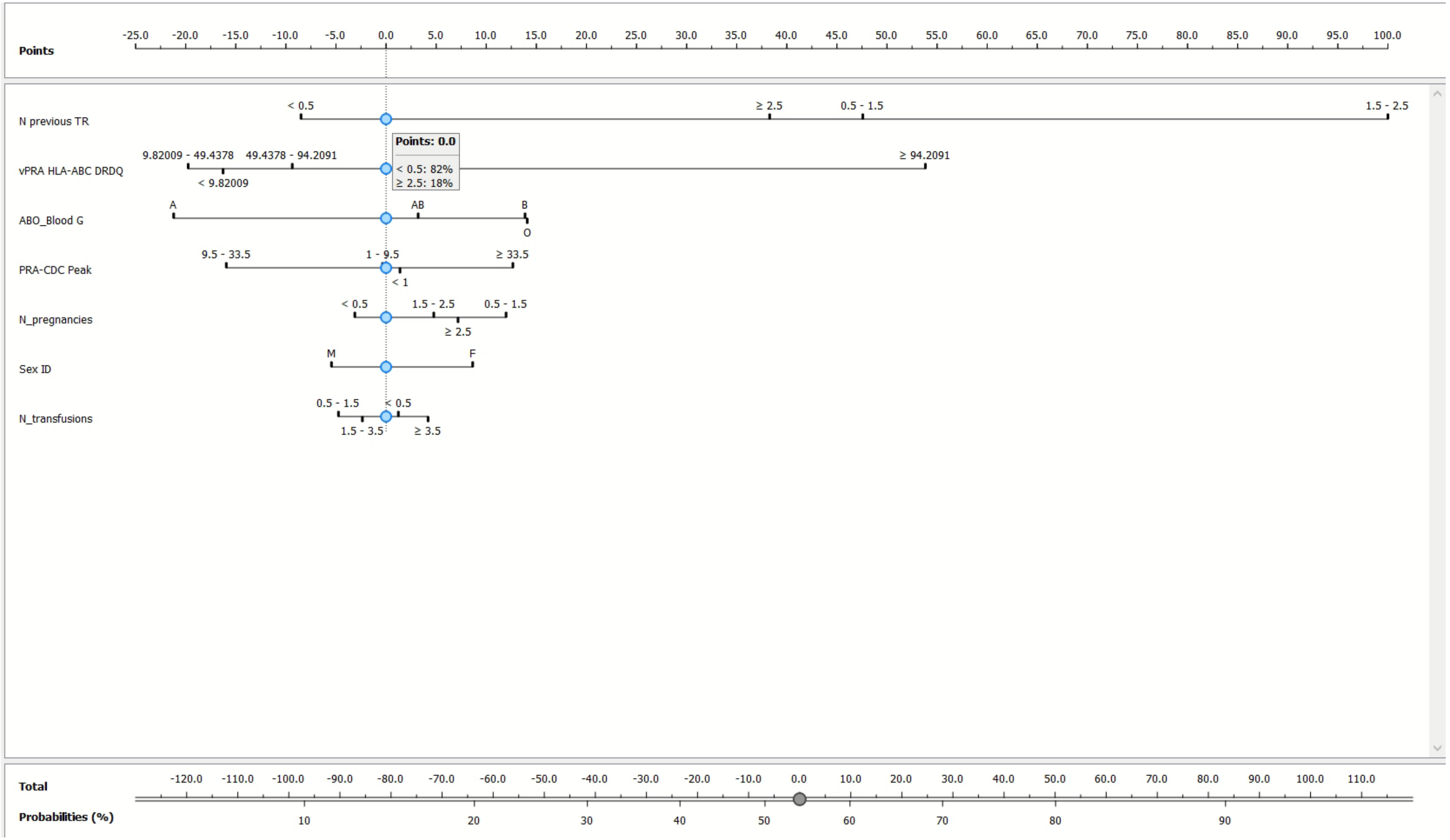
Who stays in the waiting list vs transplanted patients?

Based on vPRA HLA-ABC DRDQ



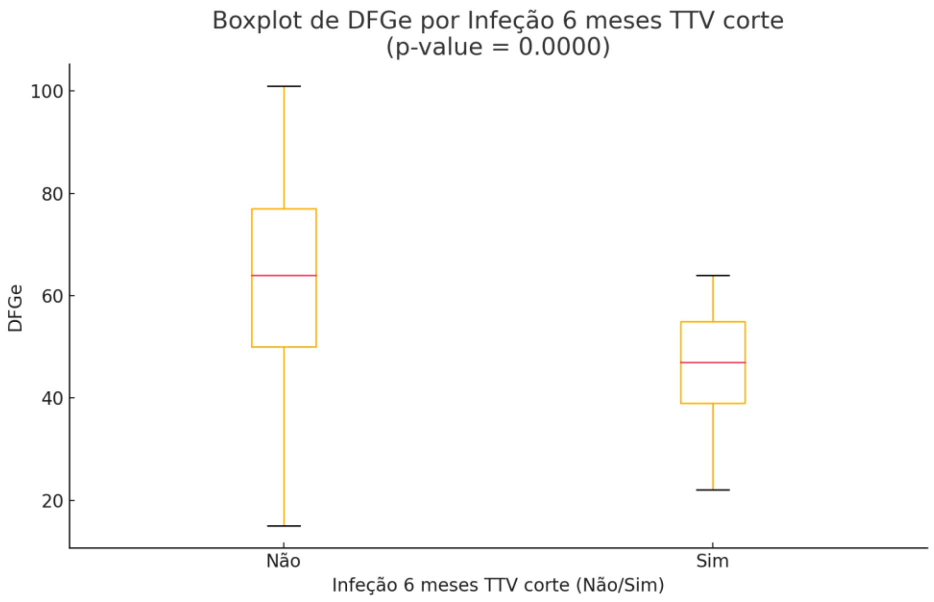
Who stays in the waiting list vs transplanted patients?

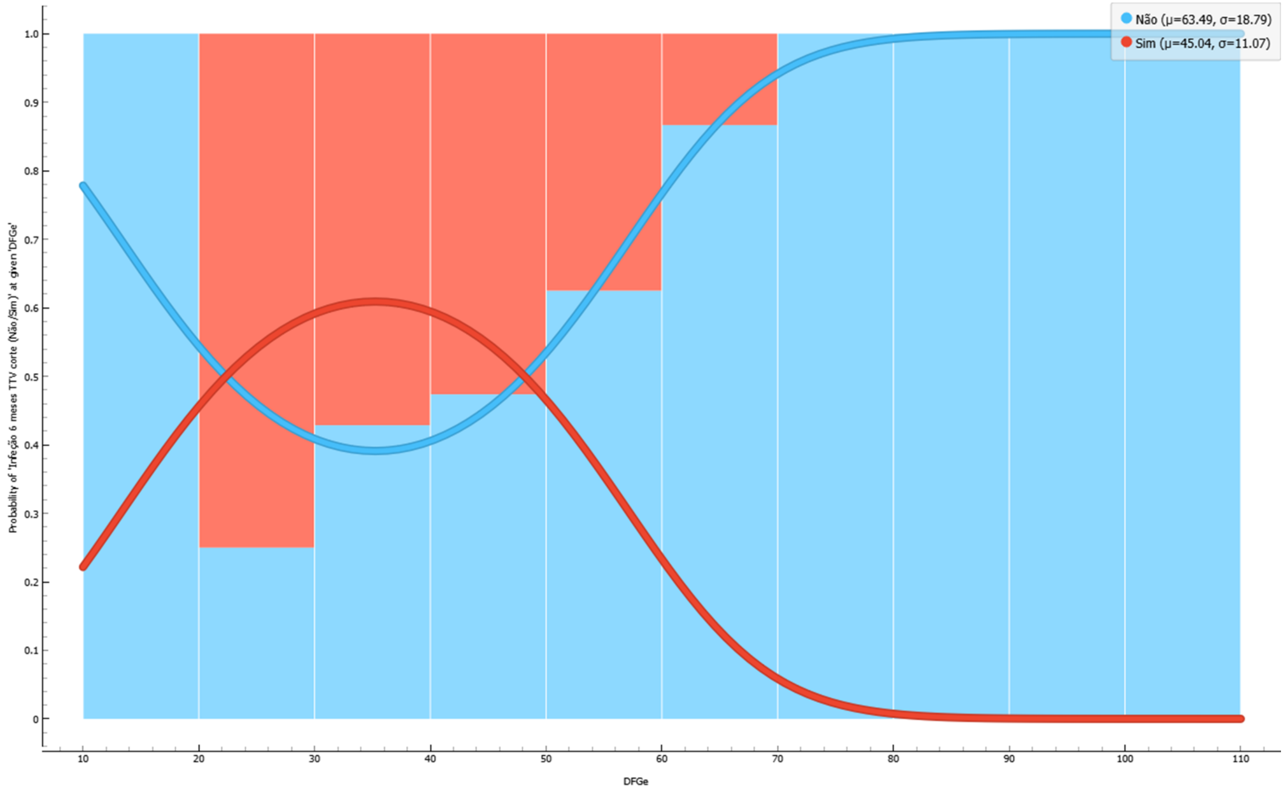
Near future



RESULTS

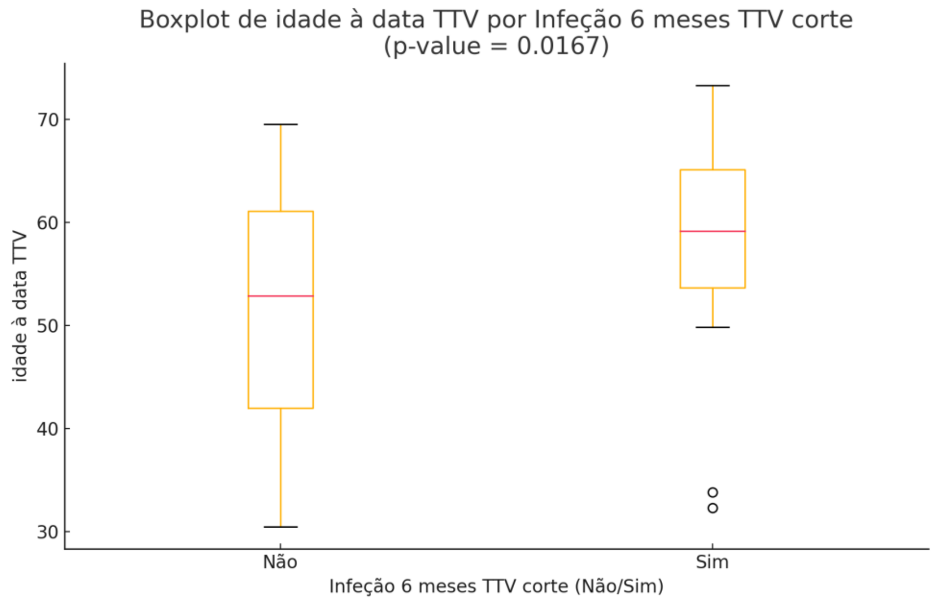
Risk of infection according to eGFR at TTV cohort analysis

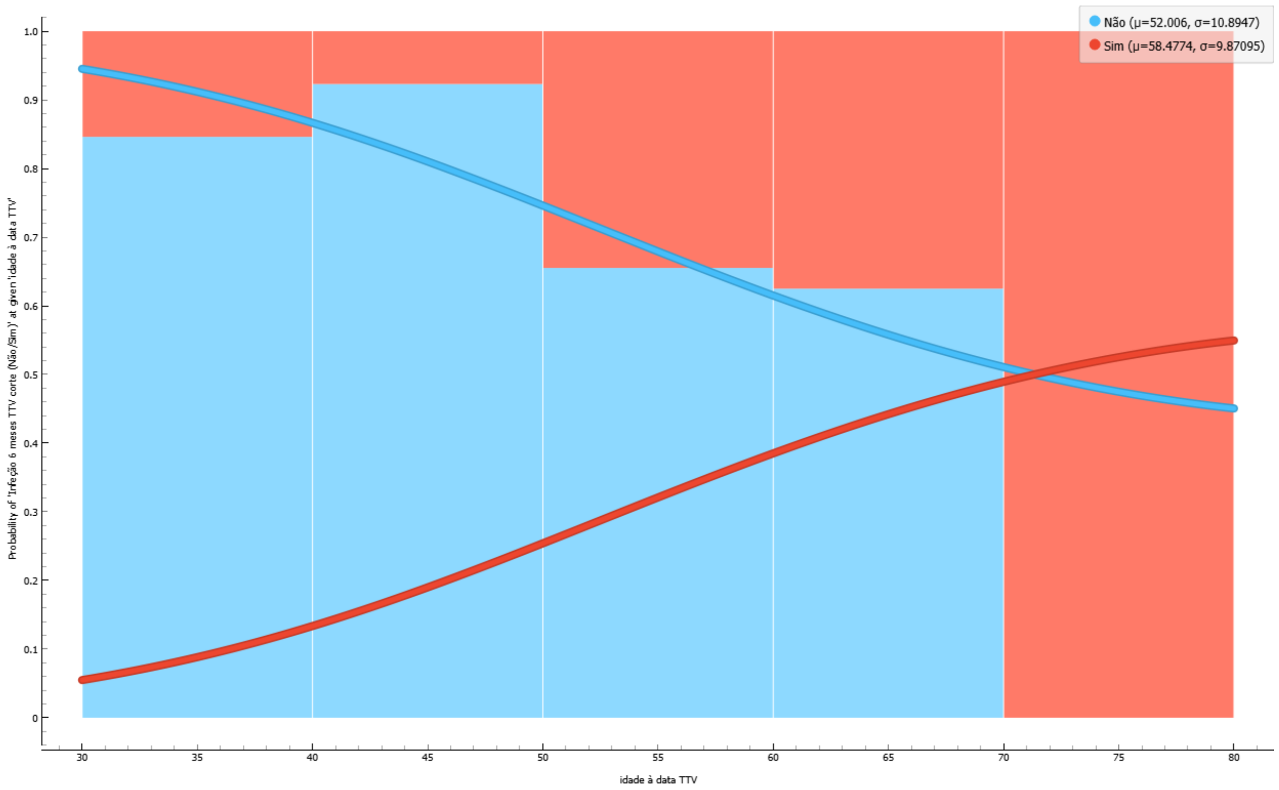




RESULTS

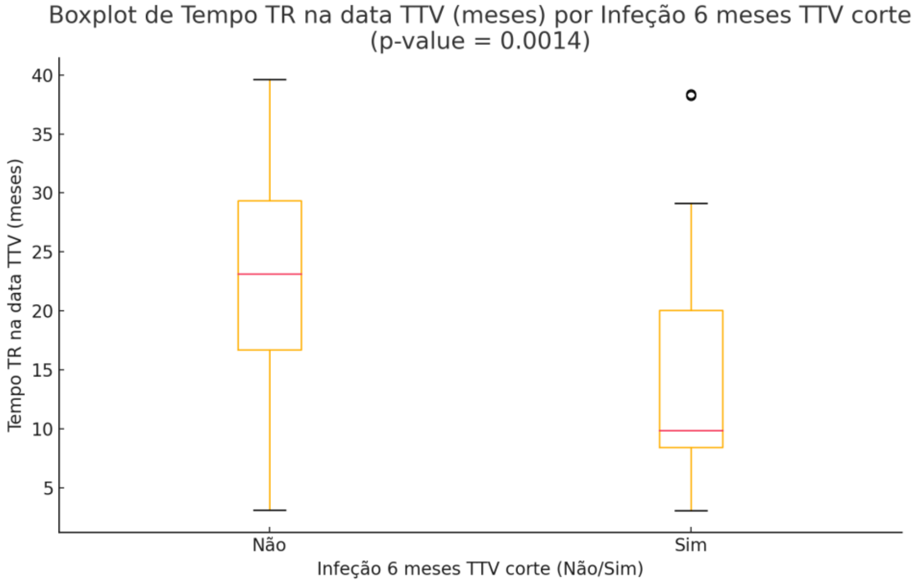
Risk of infection according to age at TTV cohort analysis

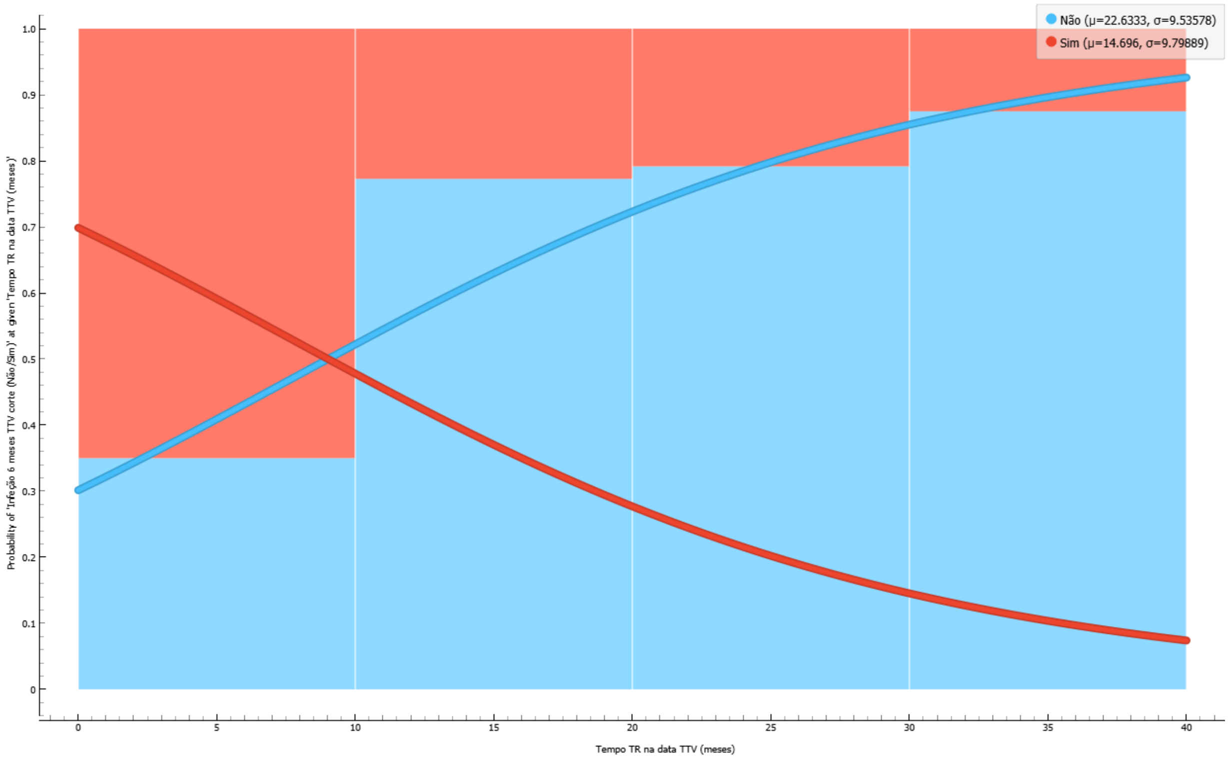




RESULTS

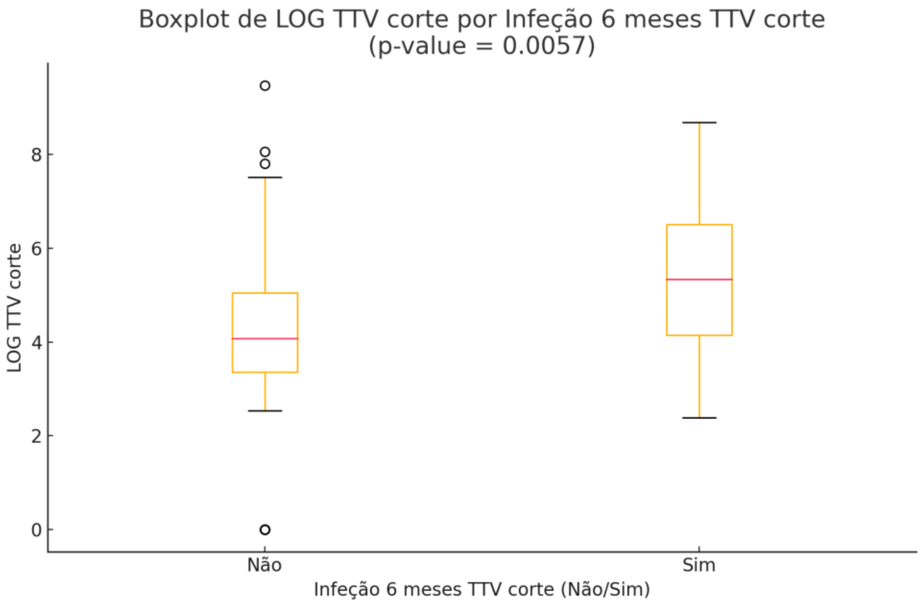
Risk of infection according to time after KT TTV cohort analysis

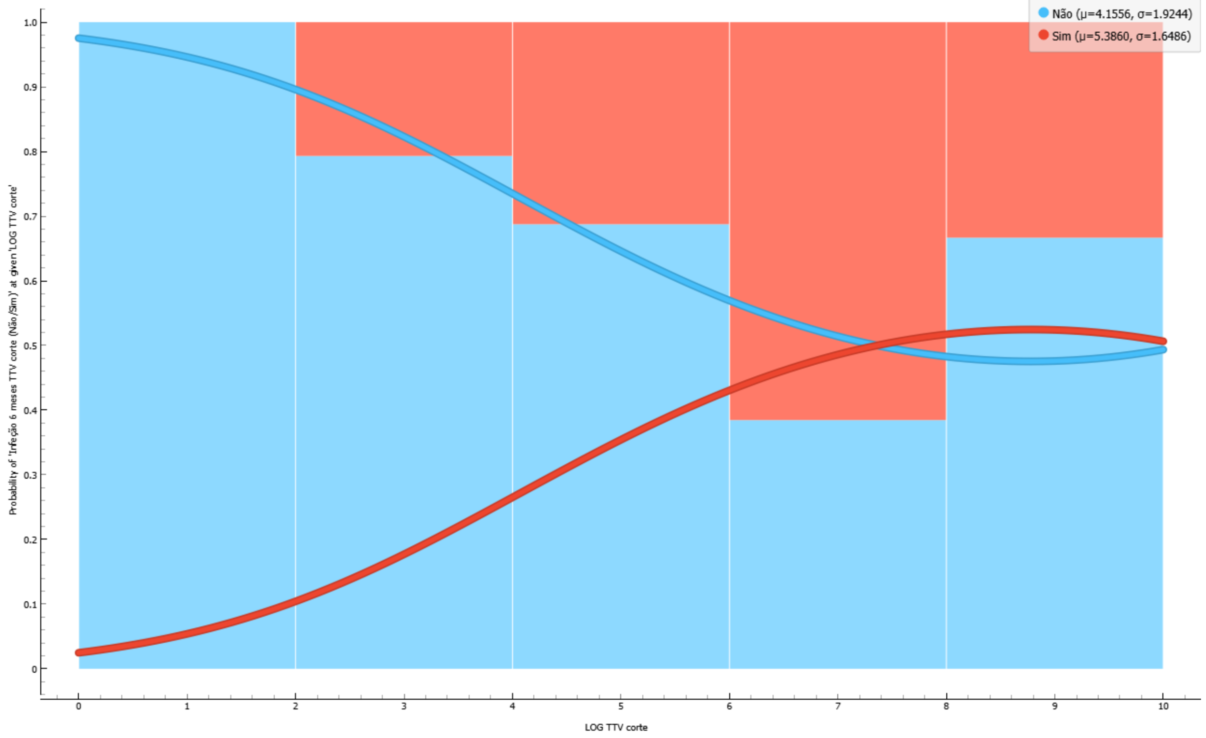




RESULTS

Risk of infection according to TTV viral load at cohort analysis





RESULTS

Proposed plasma TTV load cutoff values determined in KT recipients for the risk prediction of allograft infection (total and subgroup analysis)

Predictor	AUC	Optimal Threshold	Sensitivity	Specificity	PPV	NPV
Log TTV before KT	0,47	5,46	0,12	0,95	0,50	0,71
LOG TTV cohort	0,69	5,16	0,60	0,81	0,58	0,82
Ratio before/cohort	0,38	1,17	0,14	0,91	0,43	0,68

Timing (meses)	Predictor	AUC	Optimal Threshold	Sensitivity	Specificity	PPV	NPV	Infections (cumulative)	No infections (cumulative)
18	LOG TTV cohort	0,566666667	4,96	0,733333	0,533333	0,611111	0,666667	15	15
24	LOG TTV cohort	0,647777778	4,96	0,722222	0,68	0,619048	0,772727	18	25
30	LOG TTV cohort	0,658571429	4,96	0,65	0,714286	0,565217	0,78125	20	35
36	LOG TTV cohort	0,663690476	5,16	0,6	0,761905	0,545455	0,8	20	42
42	LOG TTV cohort	0,642929293	5,16	0,590909	0,777778	0,565217	0,795455	22	45
48	LOG TTV cohort	0,625	5,16	0,590909	0,74	0,5	0,804348	22	50

AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value.

RESULTS Machine learning: Selected variables

Model	AUC	CA	F1	Prec	Recall	MCC	Spec
Logistic Regression	0.797	0.805	0.805	0.805	0.805	0.540	0.735

		#	Info. gain
1	N DFGe		0.236
2	N Tempo TR na data TTV (meses)		0.123
3	N LOG TTV corte		0.091
4	N Ratio pre_corte		0.079

Predicted

Actual	Não	Sim	Σ
Não	49	8	57
Sim	8	17	25
Σ	57	25	82

RESULTS Machine learning: Selected variables

Target Class: Infection

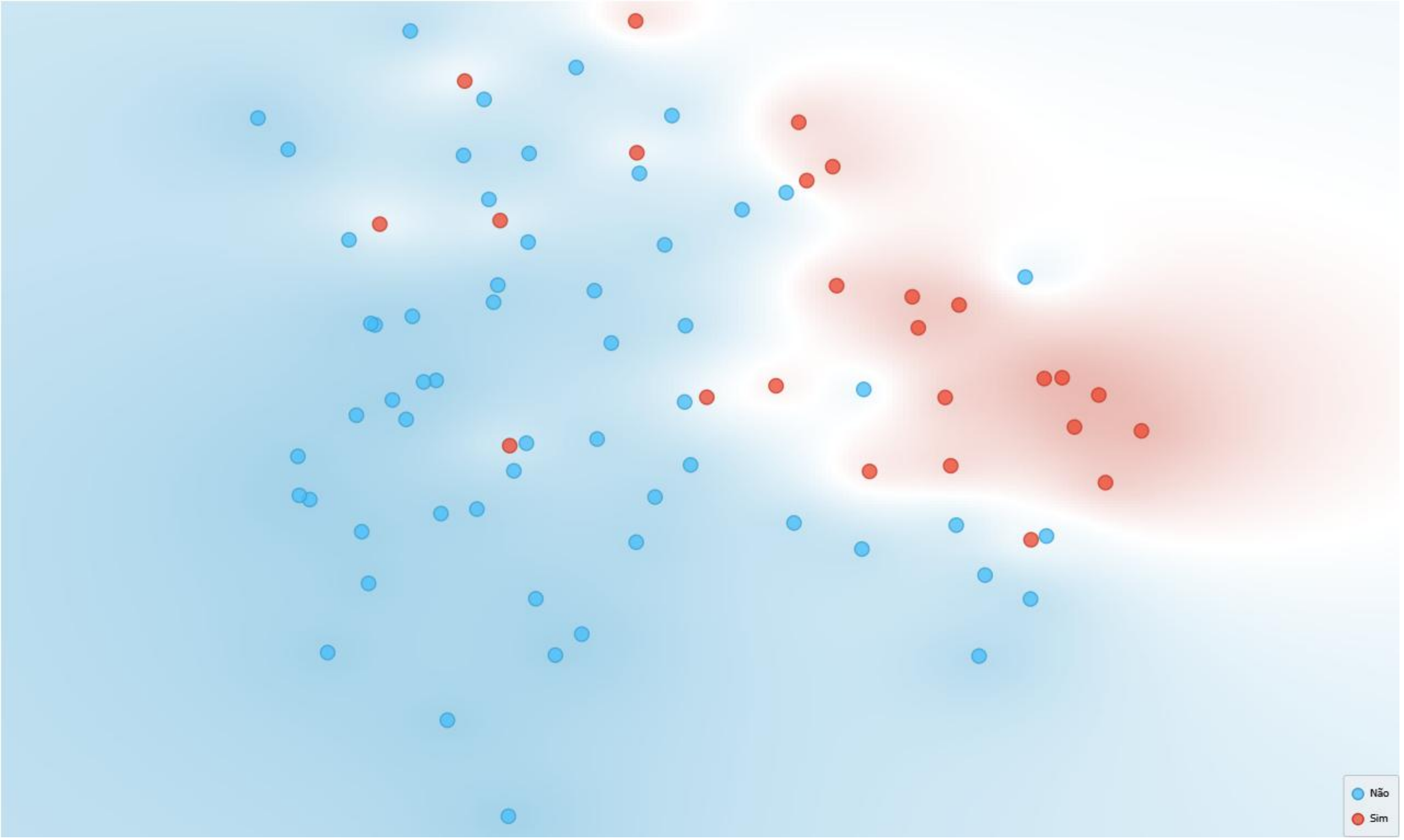
Model	CA	F1	Prec	Recall	MCC	Spec.
Logistic Regression	0.805	0.680	0.680	0.680	0.540	0.860

Target Class: No infection

Model	CA	F1	Prec	Recall	MCC	Spec.
Logistic Regression	0.805	0.860	0.860	0.860	0.540	0.680

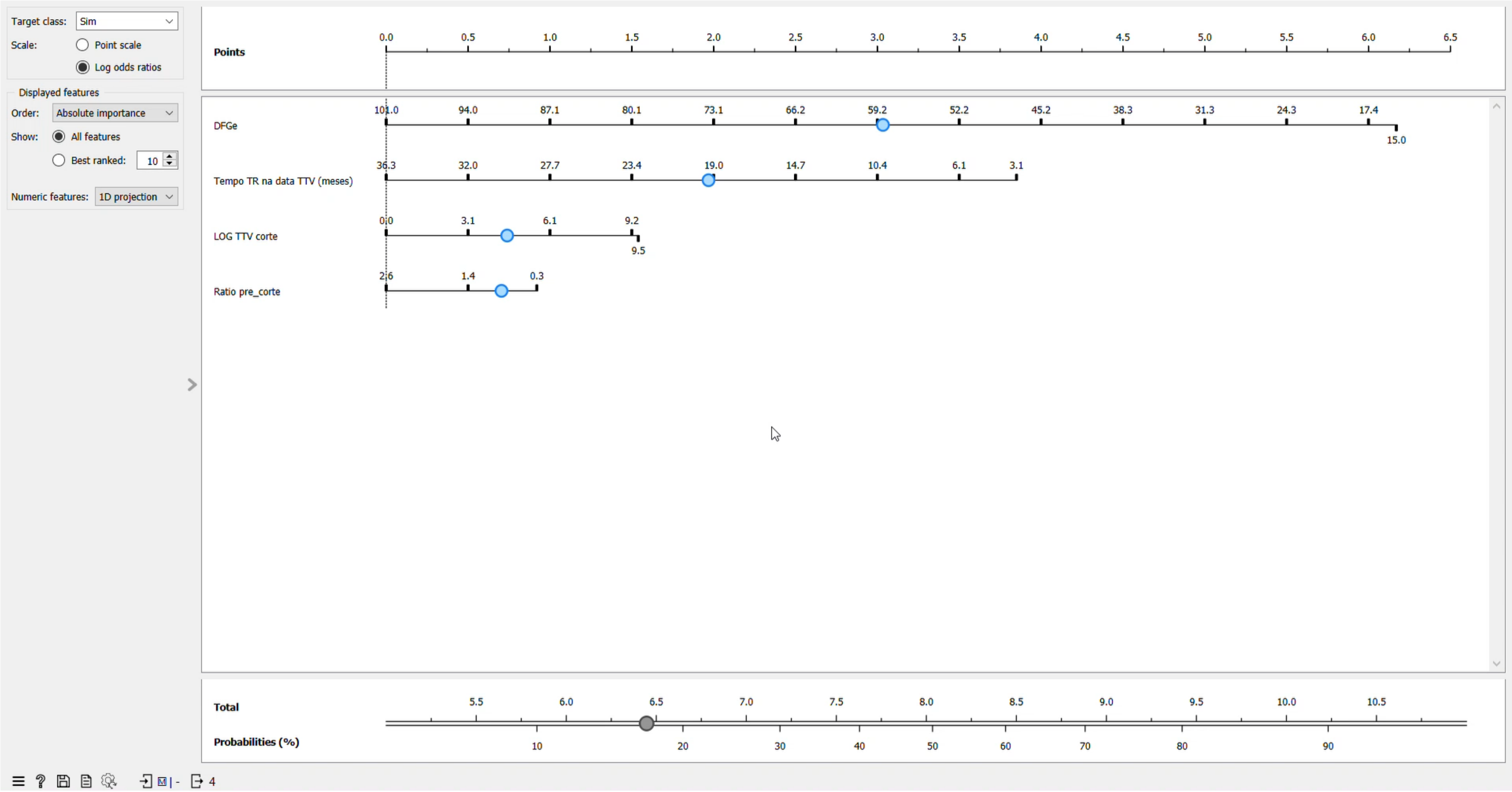
RESULTS

Machine learning: Selected variables



RESULTS

Nomogram based on logistic regression by machine learning



Disfunction in Reno-pancreatic transplant

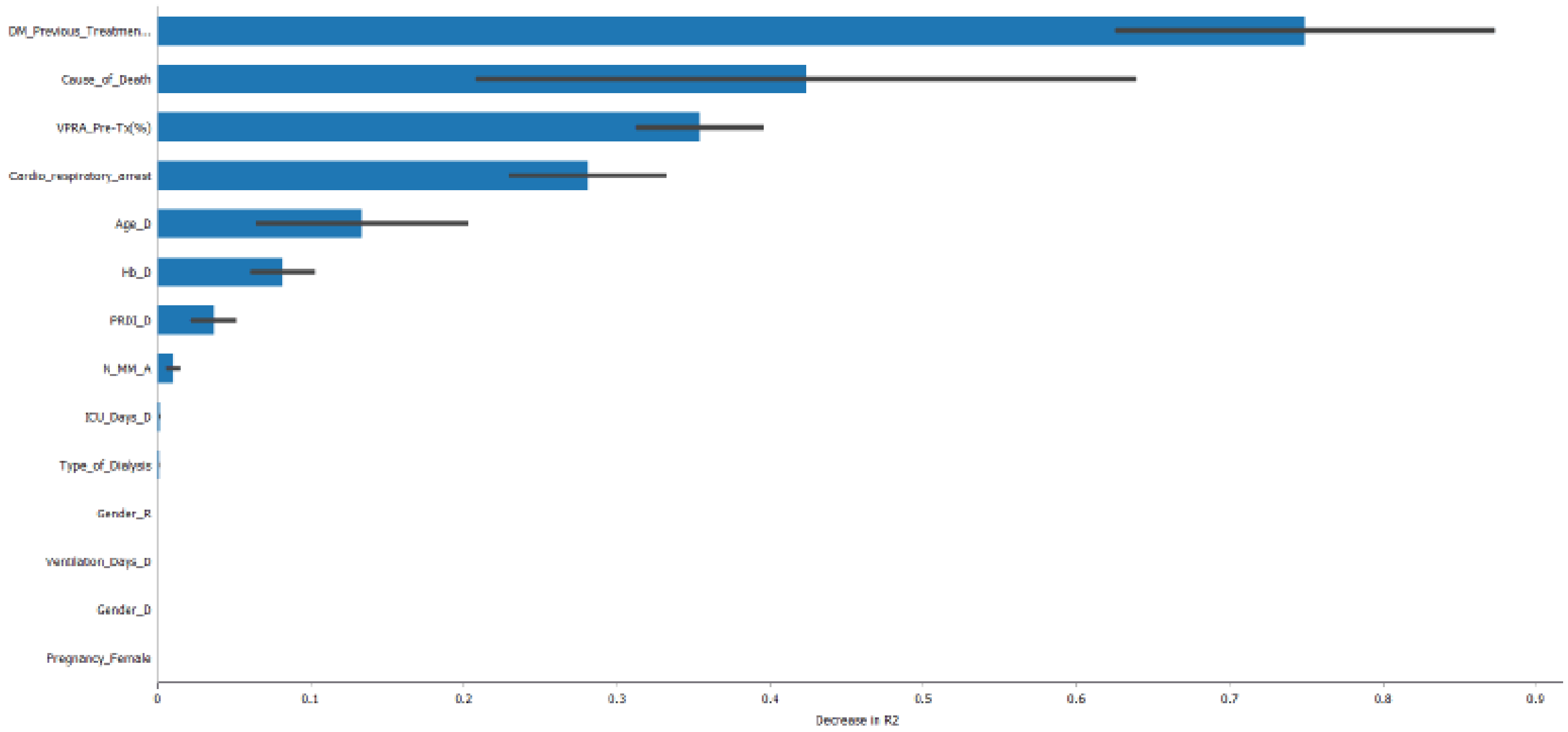


Figure 4a: Feature importance as per tree model.

Predicting Function Delay with a Machine Learning Model:
Improve the Long-term Survival of Pancreatic Grafts (Vigia, 2022)

Disfunction in Reno-pancreatic transplant

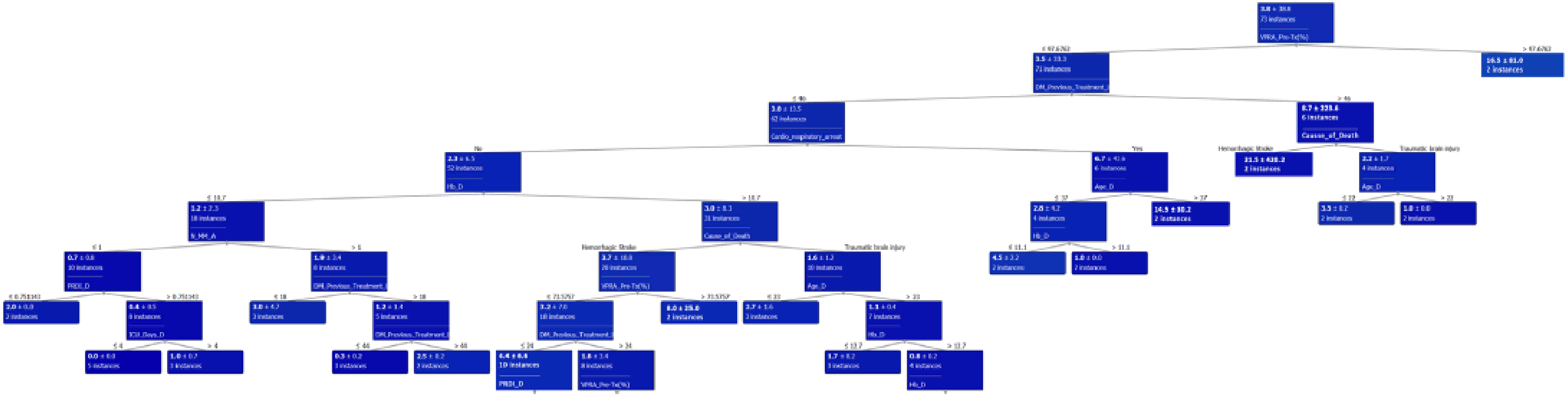


Figure 2: Tree model.

Predicting Function Delay with a Machine Learning Model: Improve the Long-term Survival of Pancreatic Grafts (Vigia, 2022)

Disfunction in Reno-pancreatic transplant

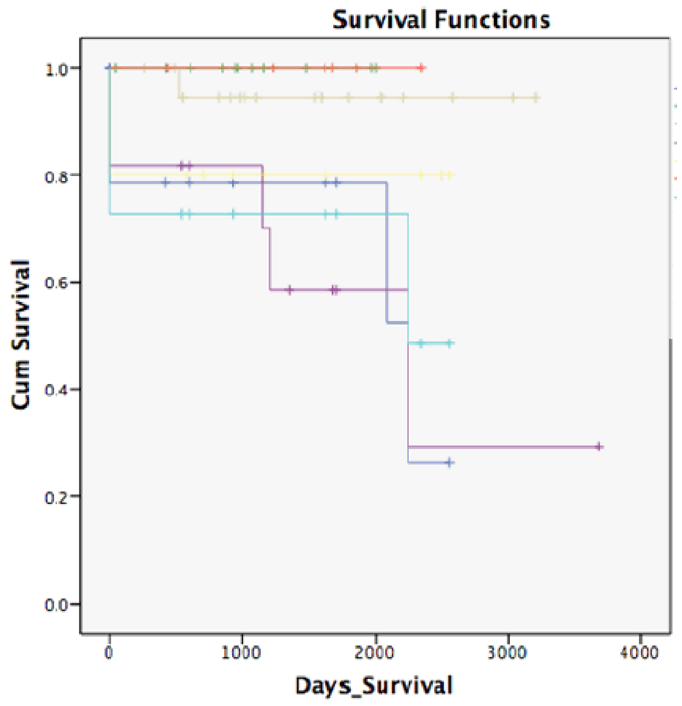


Figure 3: Graft survival.
Note: Where dark blue=discharge with insulin; green=never need insulin; dark yellow=1 day insulin; pink=2 days insulin; lite yellow=3 days insulin; red=4 days insulin; sky blue=stop and start

Predicting Function Delay with a Machine Learning Model: Improve the Long-term Survival of Pancreatic Grafts (Vigia, 2022)

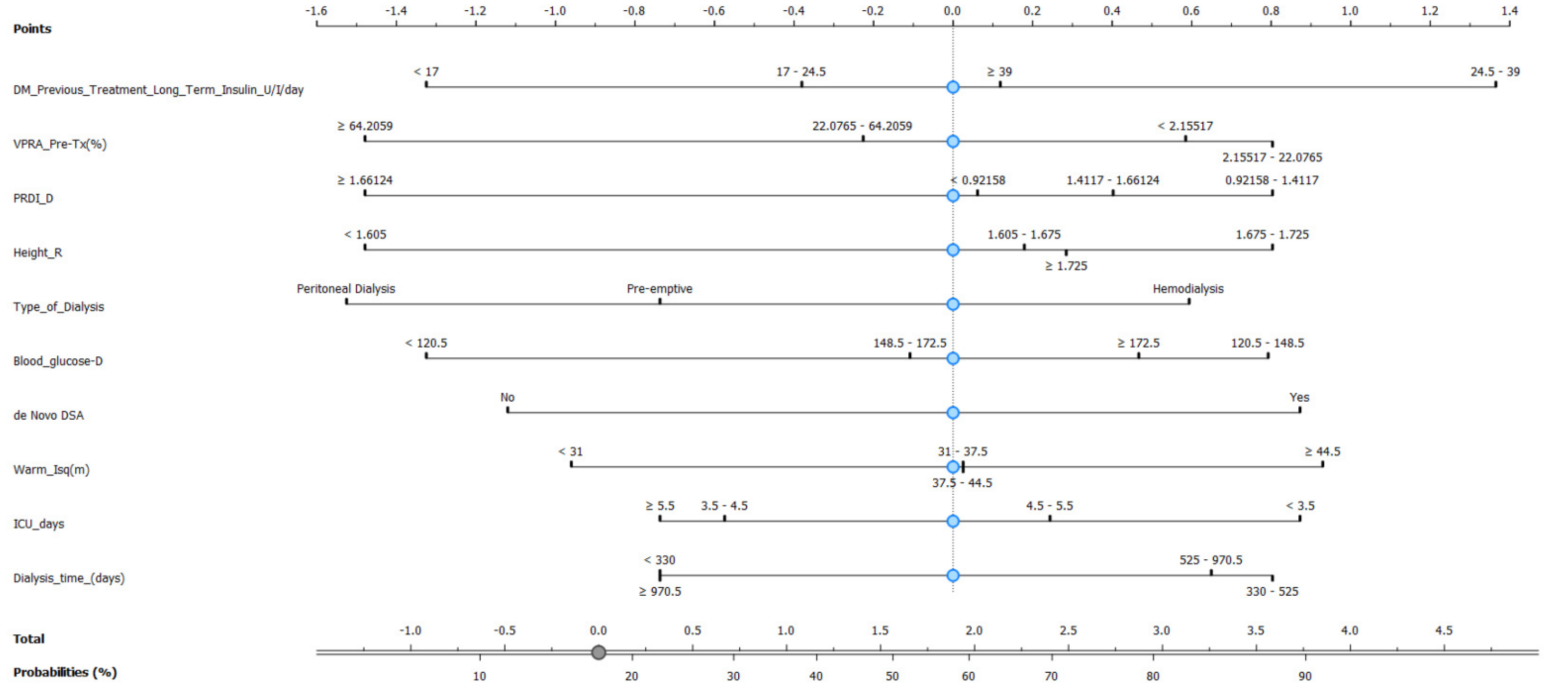
Disfunction in Reno-pancreatic transplant

A

		Predicted No	Predicted Yes	∑
Actual	No	93.9 %	6.1 %	66
	Yes	7.7 %	92.3 %	13
	∑	63	16	79

B

		Predicted No	Predicted Yes	∑
Actual	No	84.8 %	15.2 %	66
	Yes	0.0 %	100.0 %	13
	∑	56	23	79



Pancreas Rejection in the Artificial Intelligence Era: New Tool for Signal Patients at Risk (Vigia, 2023)

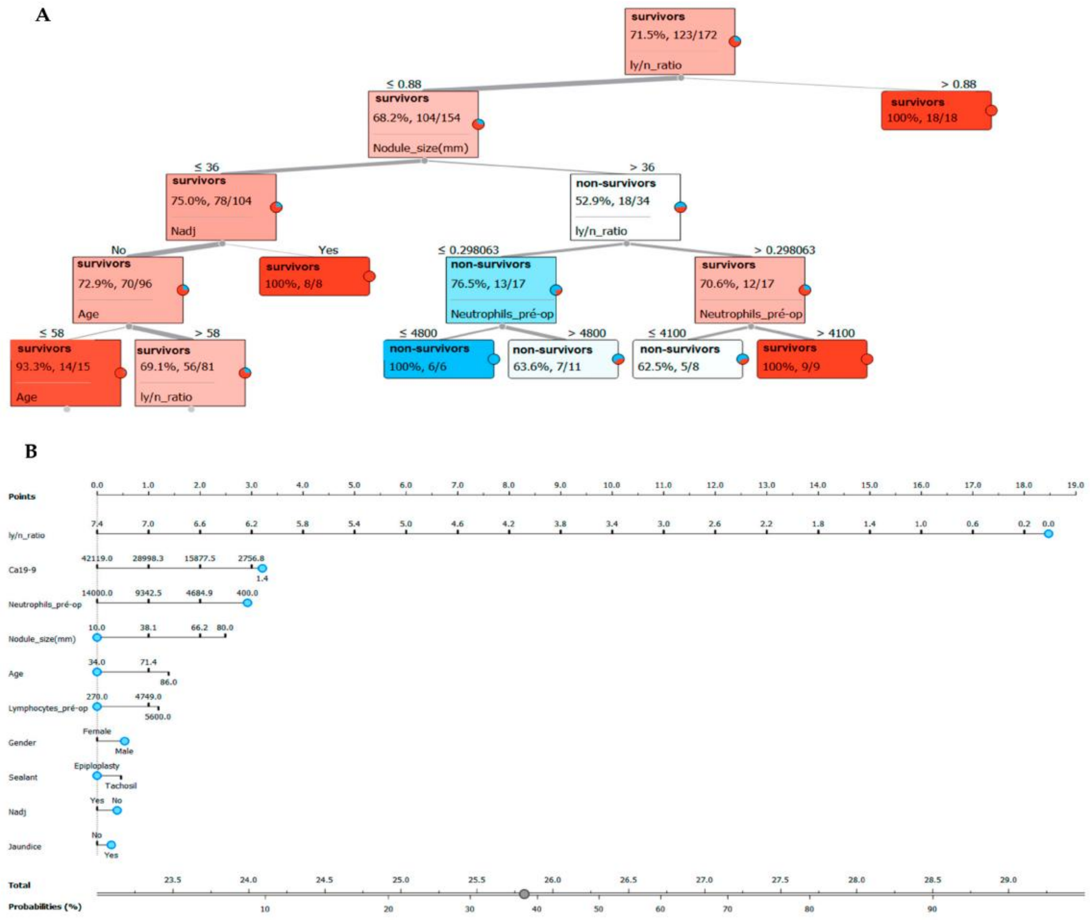
"Hmm...yes, I'm afraid it *is* a sign of cancer."

Oncology

Pancreatic Tumor

Model	AUC	CA	F1	Precision
Training Set (n=172)				
Tree	0.92	0.85	0.89	0.81
Logistic Regression	0.74	0.78	0.81	0.80
Validation Set (n=33)				
Tree	0.89	0.76	0.83	0.83
Logistic Regression	0.68	0.82	0.86	0.81

		Tree: Predicted non-survivor	Tree: Predicted survivor			Logistic Regression: Predicted non-survivor	Logistic Regression: Predicted survivor
Training Set (n=172)							
Actual	non-survivor	70.9 %	8.5 %	Actual	non-survivor	76.6 %	21.9 %
	survivor	29.1 %	91.5 %		survivor	23.8 %	78.1 %
Validation Set (n=33)							
Actual	non-survivor	55.6 %	16.7 %	Actual	non-survivor	100 %	20.0 %
	survivor	44.4 %	83.3 %		survivor	0 %	80.0 %

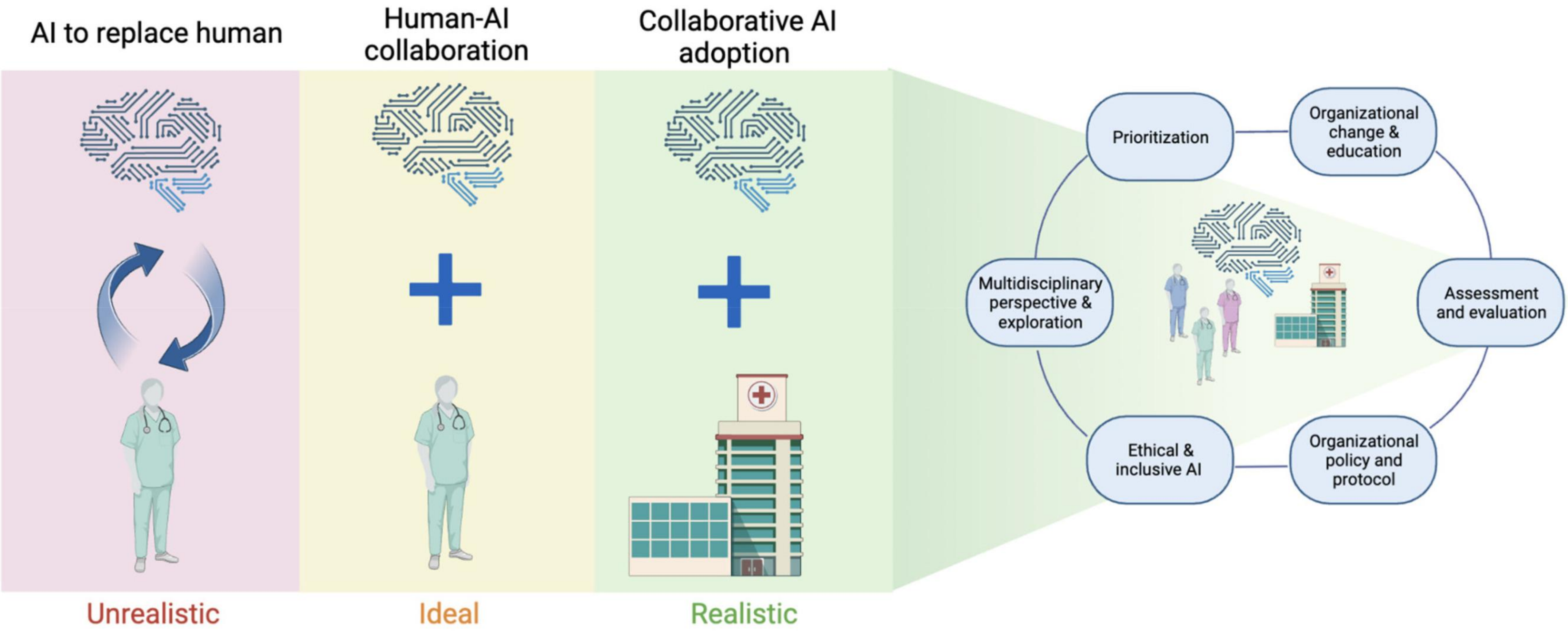


Machine Learning-Based Model Helps to Decide which Patients May Benefit from Pancreatoduodenectomy (Vigia, 2023)

HEALTHCARE

CHALLENGES

When AI is not so smart after all…


AI to replace human
Human-AI collaboration
Collaborative AI adoption
Prioritization
Organizational change & education
Multidisciplinary perspective & exploration
Assessment and evaluation
Ethical & inclusive AI
Organizational policy and protocol
Unrealistic
Ideal
Realistic

ETHICAL

Regulation

Privacy

Mitigation of Bias

Transparency

Relevance

LEGAL

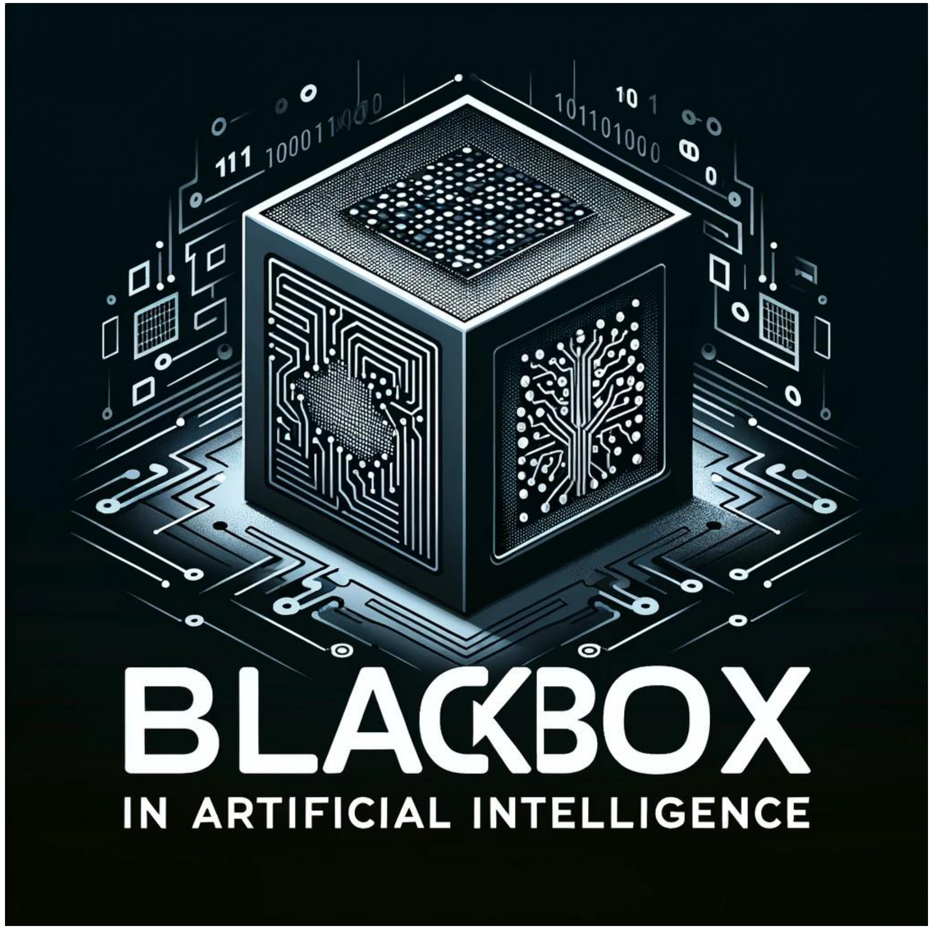
Governance

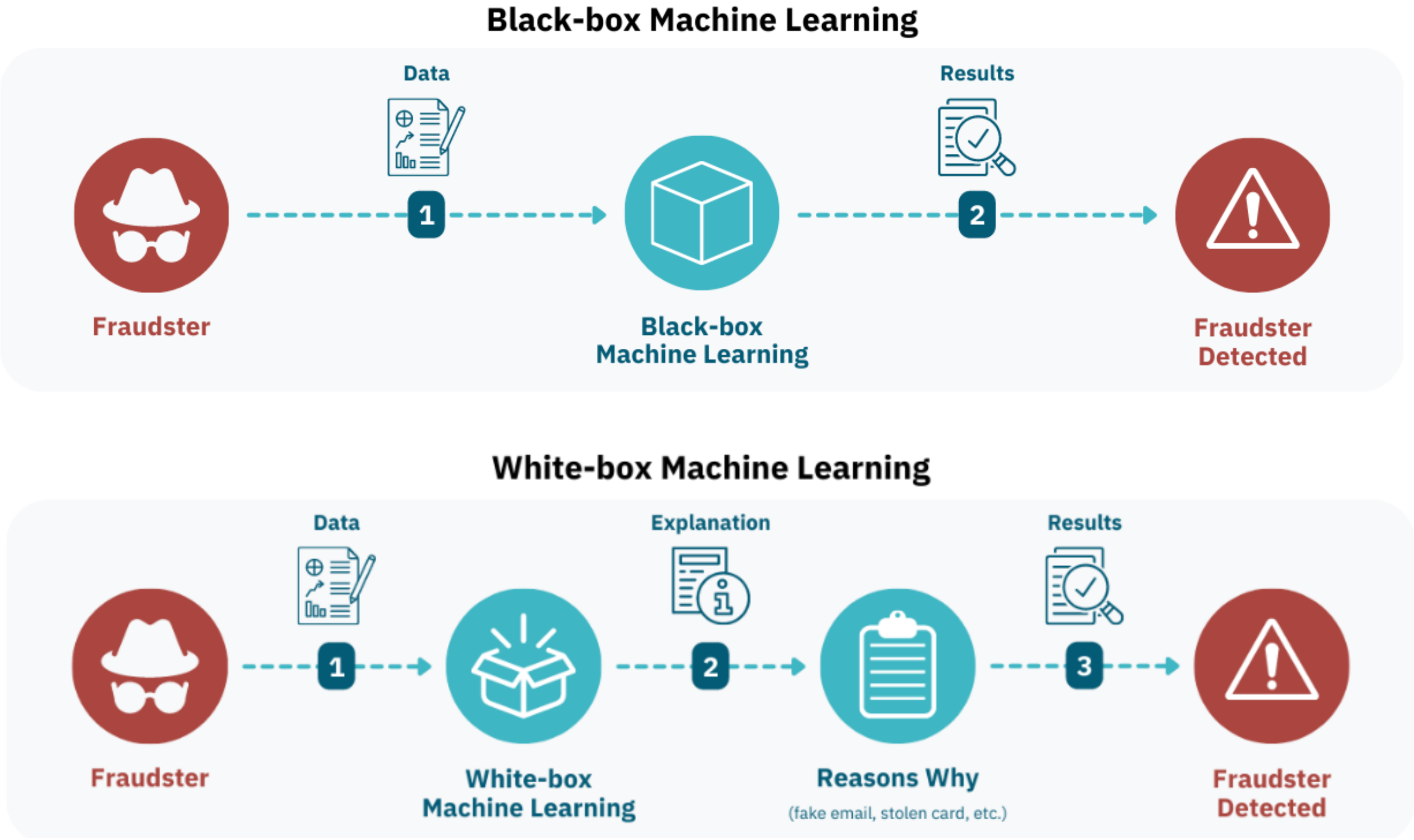
Confidentiality

Liability

Accuracy

Decision Making


BLACKBOX
IN ARTIFICIAL INTELLIGENCE


Black-box Machine Learning
Data
Results
Fraudster
1
Black-box
Machine Learning
2
Fraudster
Detected
White-box Machine Learning
Data
Explanation
Results
Fraudster
1
White-box
Machine Learning
2
Reasons Why
(fake email, stolen card, etc.)
3
Fraudster
Detected

GPT-3 (OpenAI), **Claude 1** (Anthropic), **PaLM** (Google), **LLaMA 1** (Meta).

GPT-4-Turbo (OpenAI), **Claude 3 Opus** (Anthropic), **Gemini 1.5** (Google Deepmind), **LLaMA 3 + Chain-of-Thought prompting** (Meta).

Black-Box Machine Learning	White-Box Machine Learning
Only gives an answer	Gives an answer and shows the decisions made to reach it
Does not allow humans to verify the answer or the logic	Allows humans to verify the answer and the logic
It's not possible to make real-time adjustments, and educated guesses are needed due to the opaque process	The ability to verify the logic allows for real-time adjustments
Faster, as it only needs to provide an answer	Slower, as it must provide both a result and the reasoning behind it
Works without supervision for faster and unique results	Requires human supervision to make real-time adjustments
Better for detecting anomalies or new patterns	Better for detecting historical patterns

08 Lessons Learned

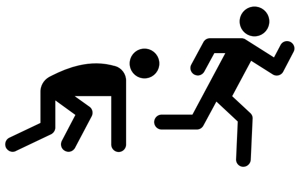
...did you learn anything?...

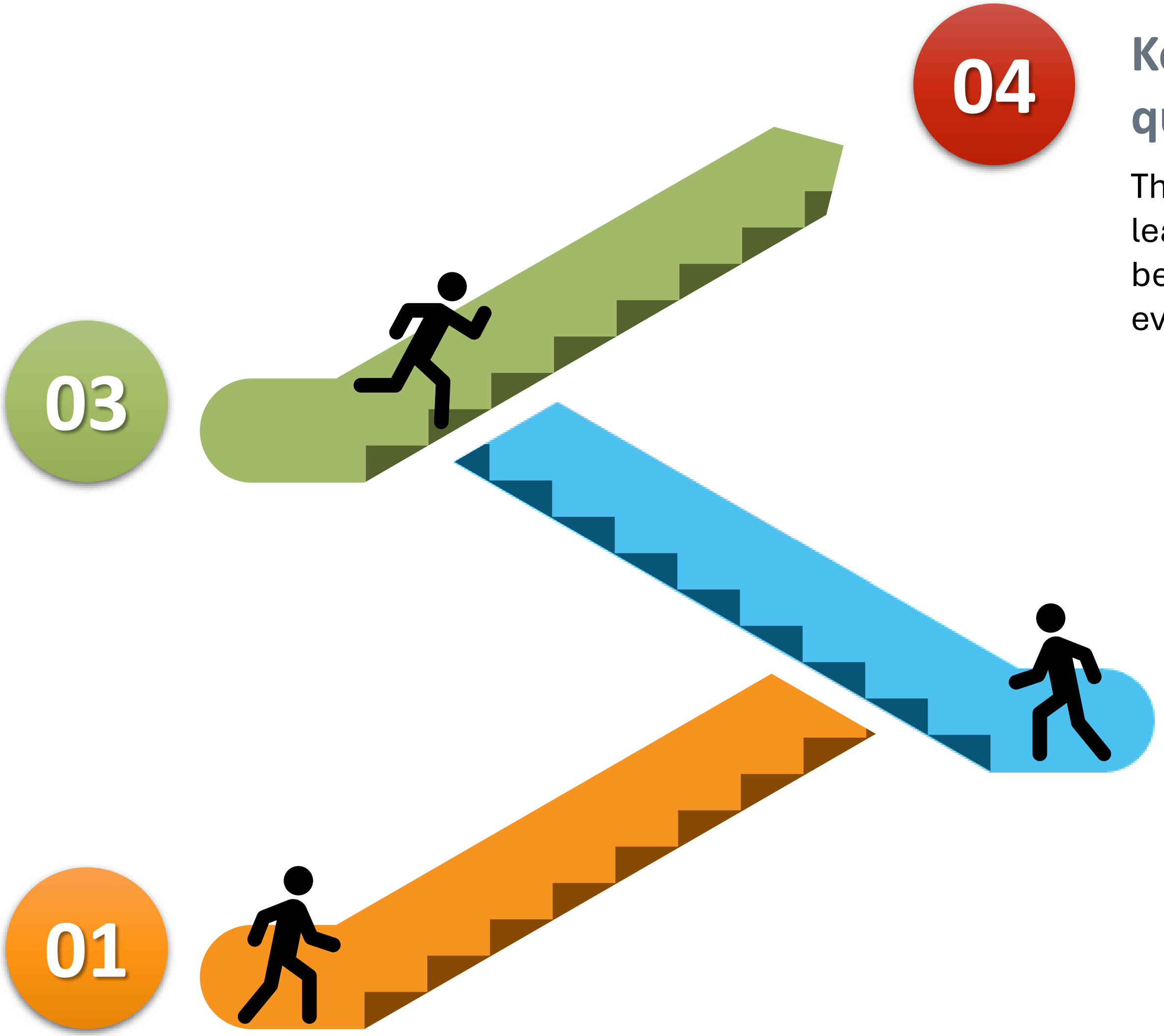
BIG Data and Healthcare

Big Data in Healthcare is vast and evolving. No matter where you start, there's room to grow, innovate, and make a difference—your background adds value.

The Climb Starts Here!

Every journey begins with effort. The path into Big Data and Healthcare is challenging, but every step builds strength, resilience, and purpose.





Keep Learning and asking questions!

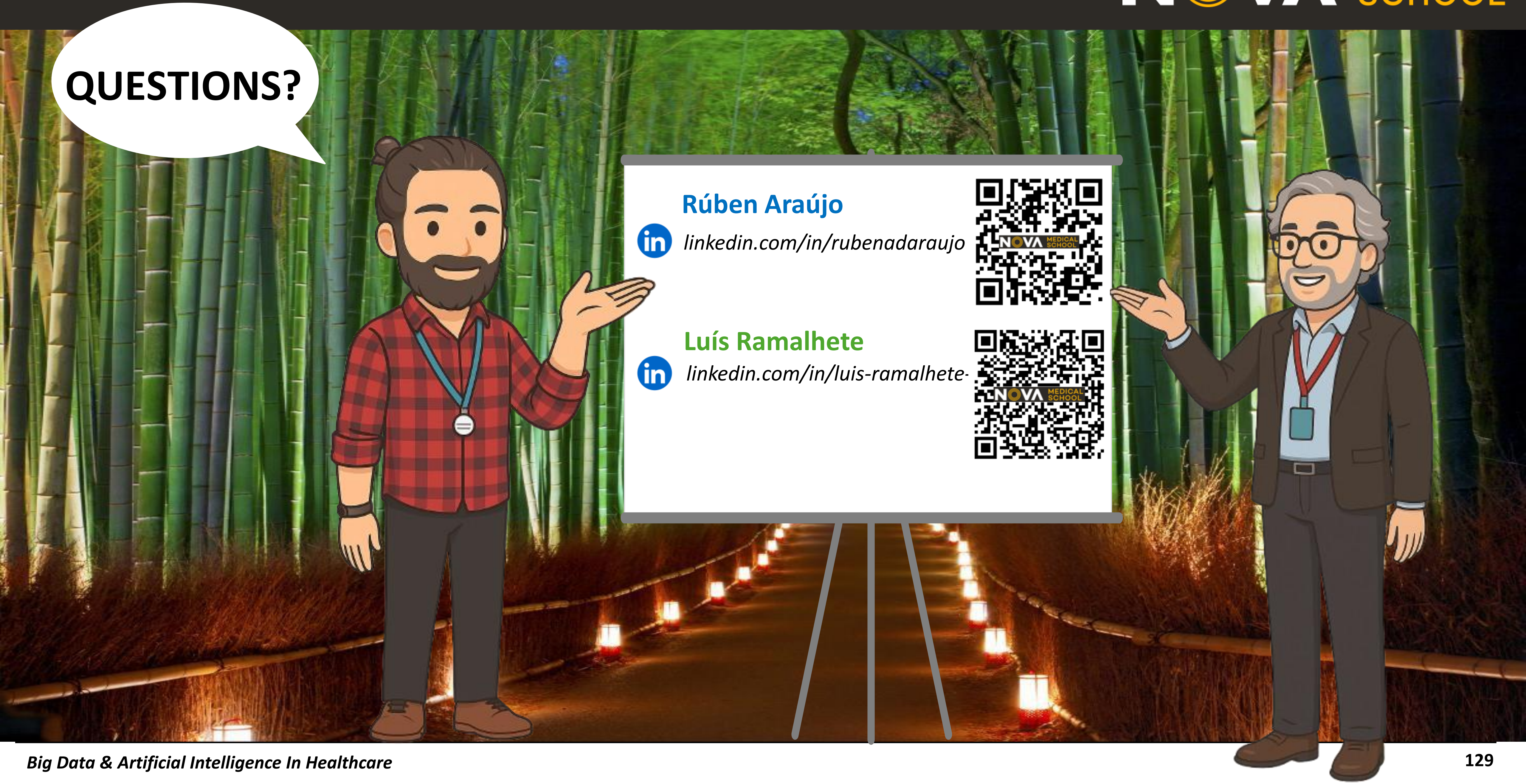
This field never stands still. Keep learning, questioning, and adapting—because curiosity is the engine behind every great discovery.

No path is the "right" one or the "only one"

Success doesn't follow one formula. Whether you come from medicine, biology, or engineering, your unique path can lead to meaningful contributions in this field.


Congratulations!
Well done! All of you!
You have survived this presentation!


THANK
YOU


QUESTIONS?
Rúben Araújo
linkedin.com/in/rubenadaraujo
Luís Ramalhete
linkedin.com/in/luis-ramalhete-

NOVA MEDICAL SCHOOL

CHRC COMPREHENSIVE HEALTH RESEARCH CENTRE

IPST Instituto Português do Sangue e da Transplantação, IP

SUMMER MEDICAL SCHOOL: LIFESTYLE

presentation on
BIG DATA & ARTIFICIAL INTELLIGENCE IN HEALTHCARE

| Guest Speakers |
Rúben Araújo – *Mechanical & Biomedical Engineer*
Luís Ramalhete – *Scientific Director of IPST Serology Laboratory*

Lisbon, 3rd July 2025