

Leveraging Electronic Health Records, Medical Knowledge, and Information Technologies for Medical Discoveries

Mina Bjelogrljic



Hôpitaux
Universitaires
Genève



**UNIVERSITÉ
DE GENÈVE**
FACULTÉ DE MÉDECINE

<https://chat.openai.com/>



Lovis Lab @HUG and UNIGE – Facmed

Medical Information Sciences (SIMED)



	André Knottnerus - EASAC (Co
	Luis Marti-Bonmati - FEAM (Co
	Rosa Castro - EASAC
	Ishita Barua - NO
	Magnus Boman - SE
	Eva Fialová - CZ
	Radko Komadina - SI
	Miikka Korja - FI
	Michal Lipschuetz - IL
	Christian Lovis - CH
	Gheorghe Ioan Mihalas - RO
	Philippe Moingeon - FR
	Alison Noble - UK
	Rui Nunes - PT
	Tania Pencheva - BG
	Renata Raidou - AT
	David Rios Insua - ES
	Daniel Rückert - UK
	Isabelle Salmon - BE
	Miroslav Samaržija - HR
	Ewout Steyerberg - NL
	Predrag Tadić - RS
	Sebastian Vollmer - DE

Lead Human machine interactions –
laetitia.gosetto@hug.ch

Lead Data semantics –
christophe.gaudet-blavignac@hug.ch

Lead Machine learning –
mina.bjelogrljic@hug.ch



UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



Hôpitaux
Universitaires
Genève



Horizon2020
European Union Funding
for Research & Innovation

HASLERSTIFTUNG



Ceres
FOUNDATION



SANOFI



AAL
PROGRAMME



Swiss National
Science Foundation



FONDATION
LEENAARDS





International Sharing of Personal Health Data for Research

April 2021

allea

All European
Academies

European Academies
easac
Science Advisory Council



Image: ALLEA / EASAC / FEAM



[HOME](#) [ABOUT US](#) [ACTIVITIES](#) [PUBLICATIONS](#) [NEWS & EVENTS](#)

NEWS & EVENTS

Federation of European Academies of Medicine

[Home](#) > [Activities](#) > [Artificial Intelligence](#) > [EASAC/FEAM joint project on AI in Healthcare](#)

[< Go back](#)

EASAC/FEAM joint project on AI in Healthcare

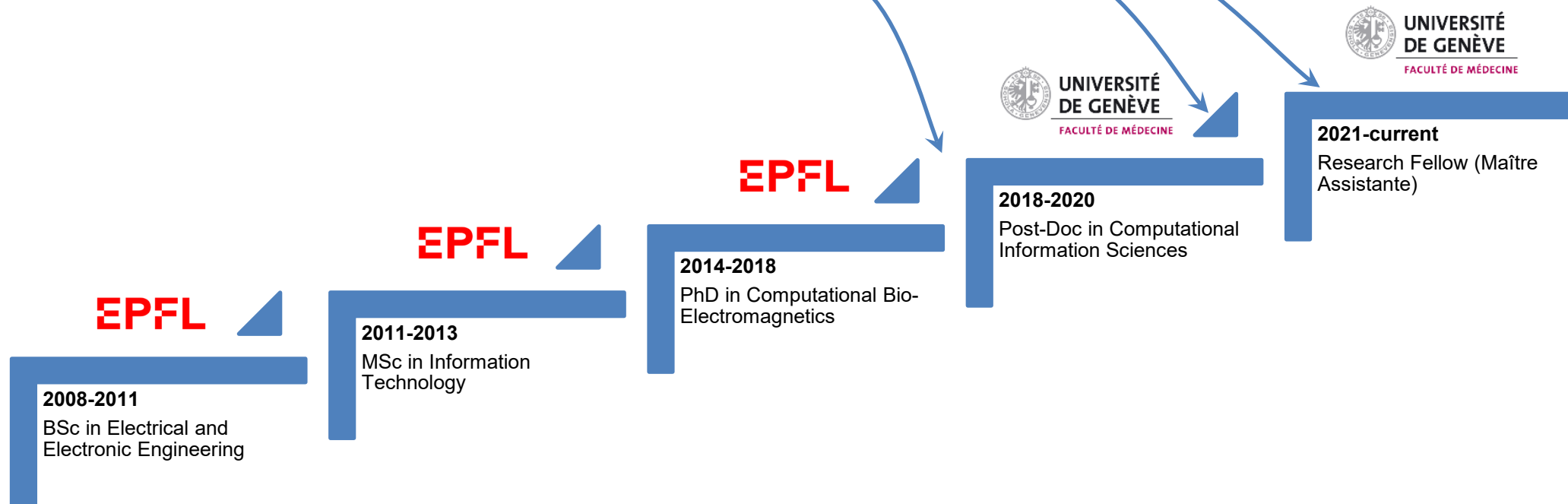
- New report on AI Transparency for Healthcare

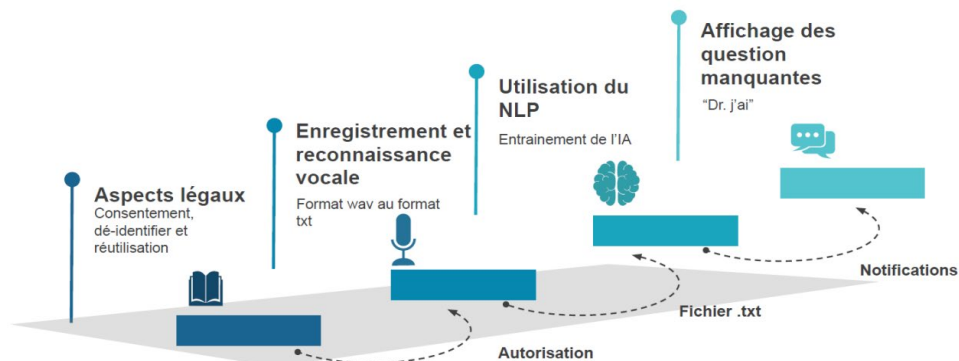
Academic Profile & Education

Mina Bjelogrljic: Academic Age: 6+ years, with experience in teaching, mentoring, and significant contributions to AI in healthcare

Academic breaks:

- ❖ Maternity (6 months 2021)
- ❖ Clinical activities – Covid (10 months 2020)
- ❖ Project Management Training (4 months 2018)





Run notebook

Introduction Python et IA

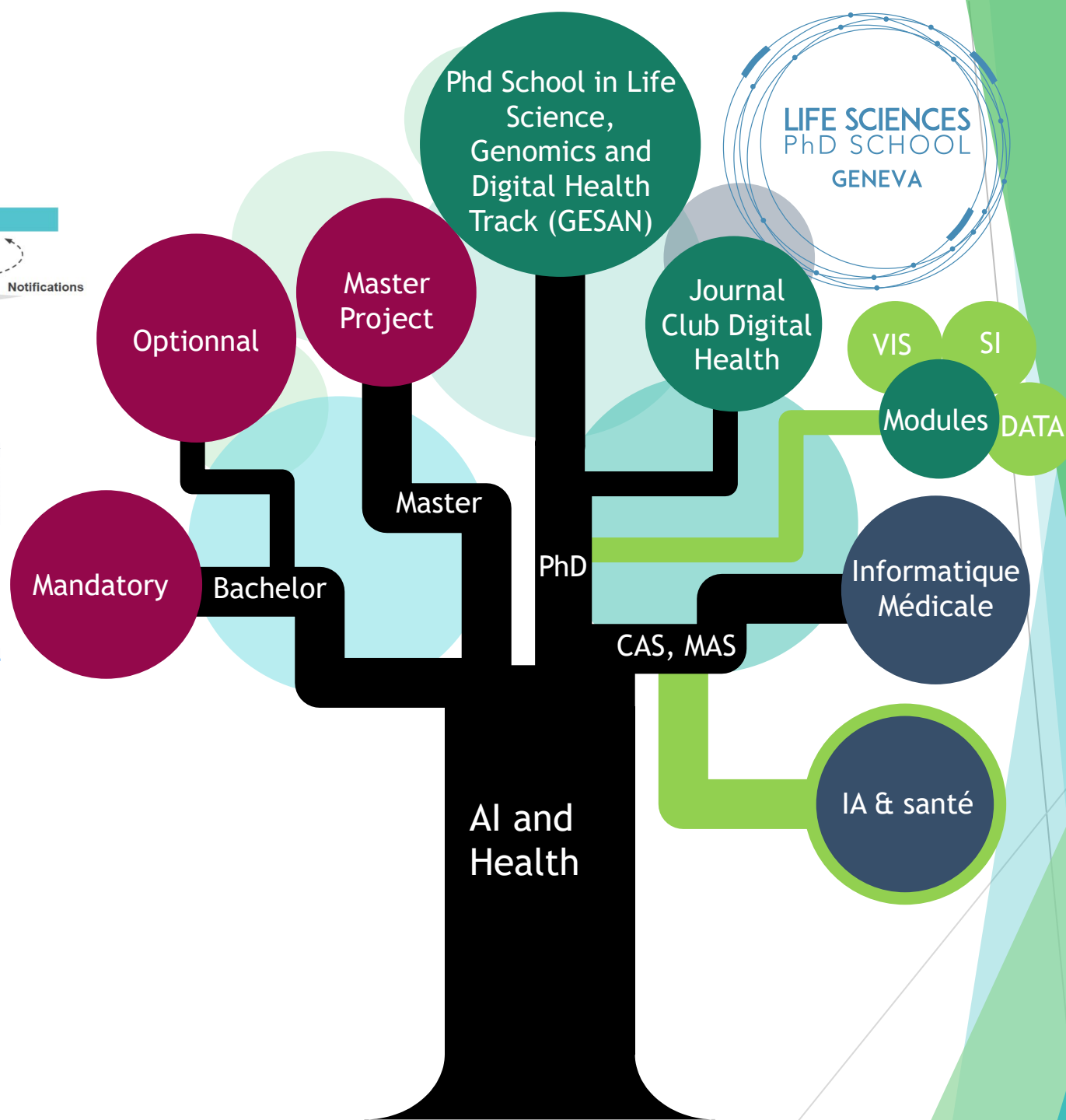
** Author: ** Mina Bjelogrić

** CO: ** IA et Science des données médicales

Press Enter or double-click on this block to edit it.

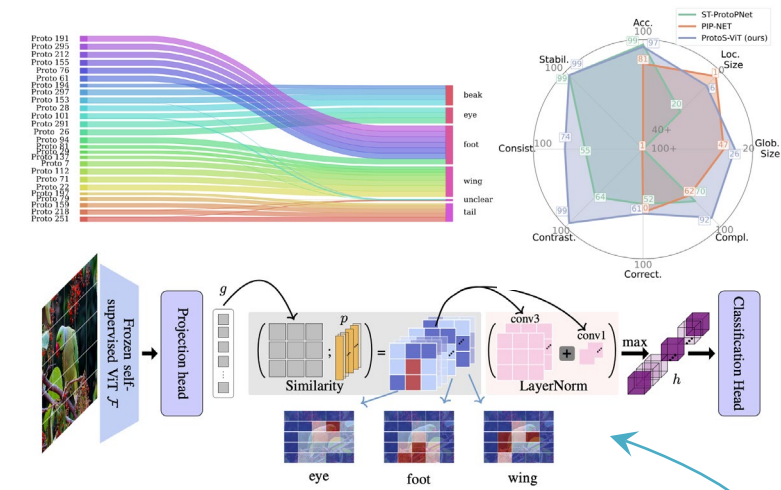
Python: les basiques

- Ensemble de règles pour dire à un ordinateur comment il doit procéder (algorithmes)
- Donne un cadre linguistique pour décrire des opérations
 - un alphabet
 - un vocabulaire
 - des règles grammaticales → une syntaxe

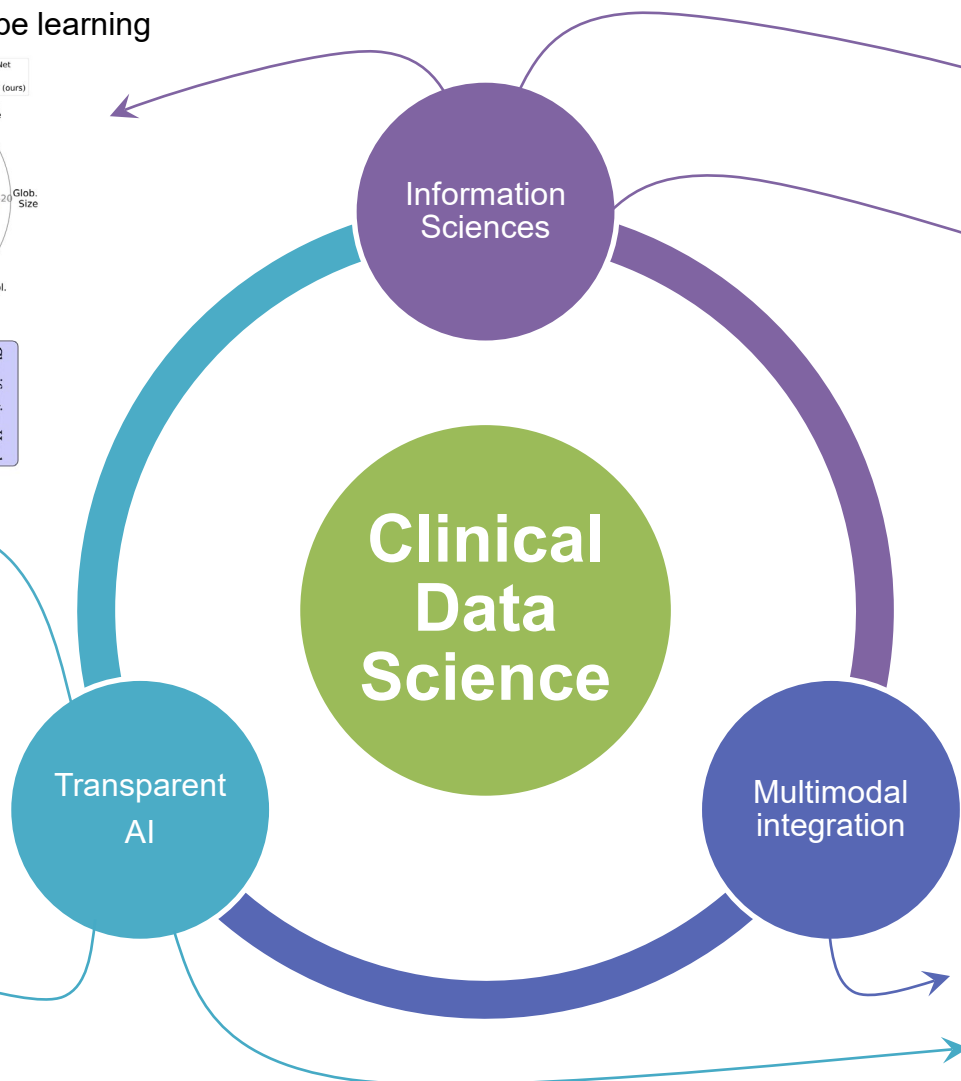
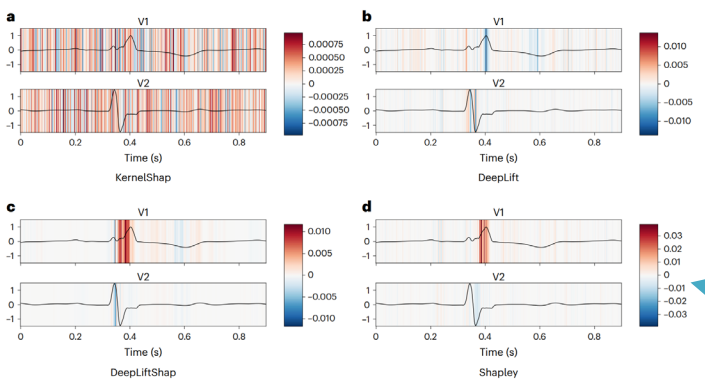


Research Focus & Expertise

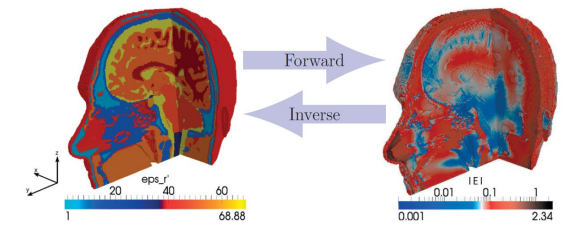
By-design explainable models via sparse prototype learning



Interpretable AI predictions



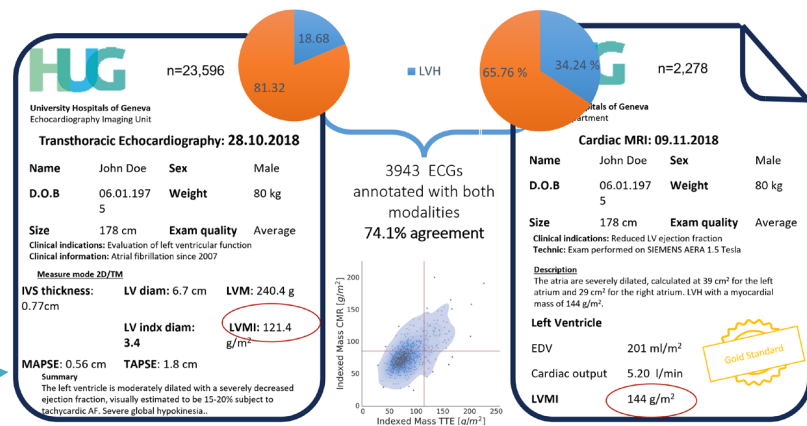
Computational electromagnetics & image/signal processing for medical applications



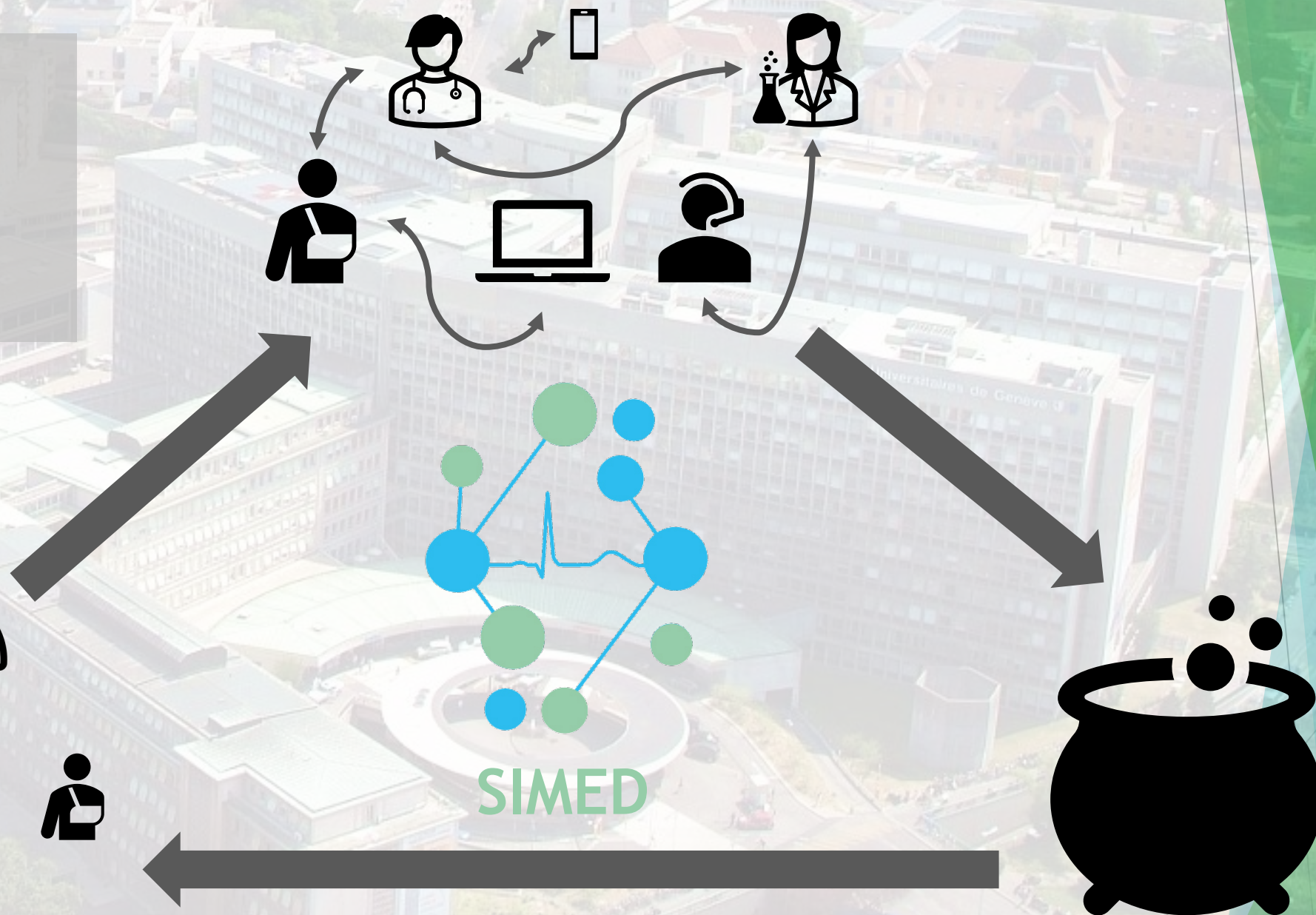
Automatic annotations & Medical entity linking



Multimodal classification models based on textual and time-series (e.g. ECG) from HUG EHRs



9 Hospitals
2 Clinics
30 Health facilities
63'481 Inpatients
3'922 Births
1'206'624 Outpatients
249'276 Patients
12'803 Collaborators



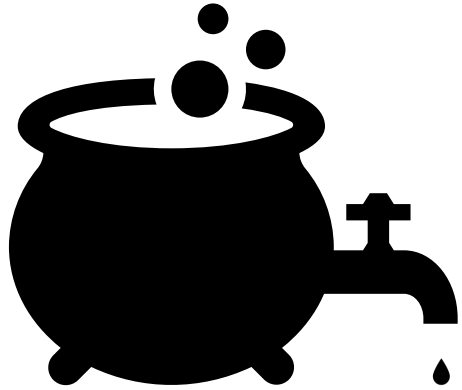
Secondary use, a challenge!



UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



Hôpitaux
Universitaires
Genève



Combien de patients covid?

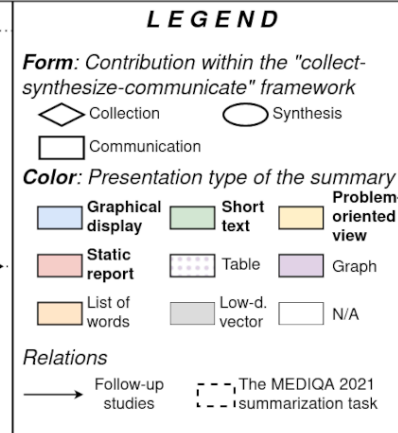
Qui a une fracture du rachis?

Combien d'enfants avec une fracture non guérie?





- Non-controlled data acquisition system
- High dimensionality
- Complex interactions



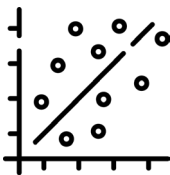
Clinical data: wide data rather than big data



Temporal



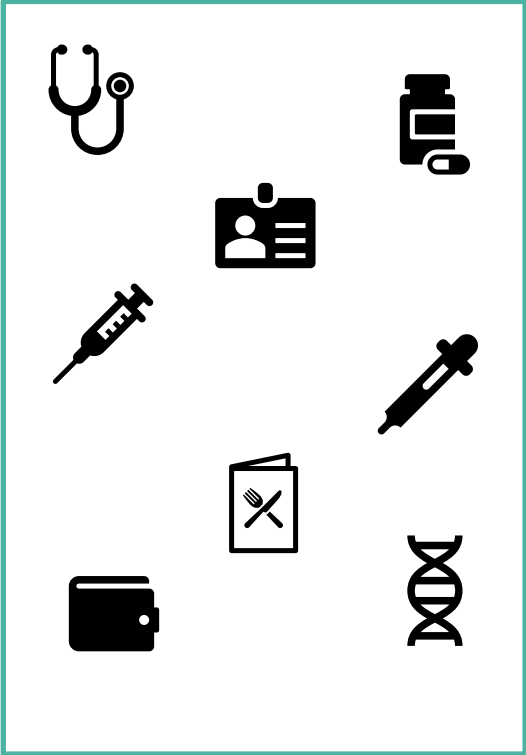
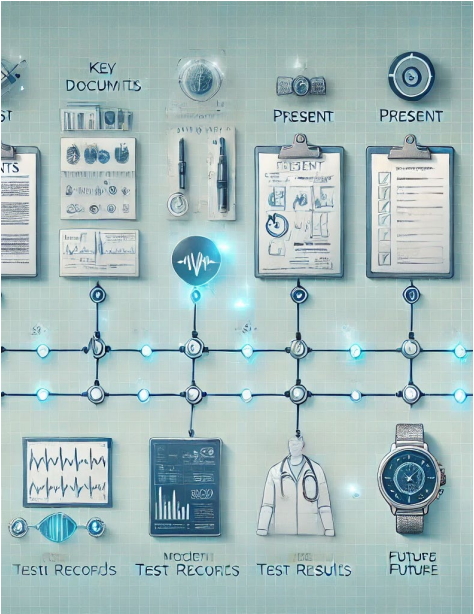
Heterogeneous



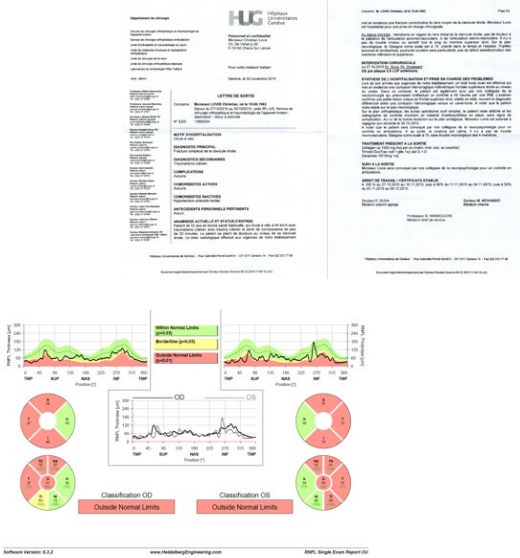
Sparse

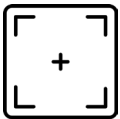


multimodal

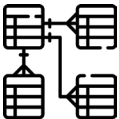


Variable	Diagnostics				Traitements			
	D1	D2	...	Dn	T1	T2	T3	Tn
Patient 1		x				x		
Patient 2								
...								
Patient n-1	x							x
Patient n				x				
Total	x	x	x	x	x	x	x	x

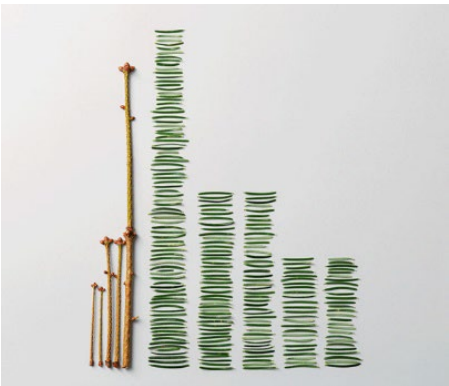




Context



Modelisation



Coding



Blanche neige endormie entourée des sept nains éplorés.

R40.2 Coma
T57.0 Effets toxiques : Arsenic et ses composés
E34.3 Insuffisance staturale, non classée ailleurs
Z63 Autres difficultés liées à l'entourage immédiat, y compris la situation familiale

Réel

Narratif

Code

Quantité d'information

Examen physique ^

> ☐ Score de Ferriman-Gallwey (Image pour évaluation)

> ☐ Score de Tanner

> ☒ Signes d'acromégalie

- ☐ Faciès acromégale :
- ☒ Macroglosie
- ☐ Augmentation de la taille des mains et des doigts
- ☐ Augmentation de la taille des pieds
- ☒ Hyperhydrose

Autres

"_id" : "PF_0672_3 " :

[0,0,1:

[0,1,0,0,1,""]

]



From data back to information

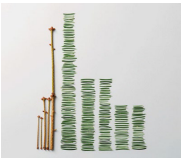
Bridge together data from medical specialties, and from different health provider systems.



Wide coverage of sources

Structured and free text

Keeping contextual information



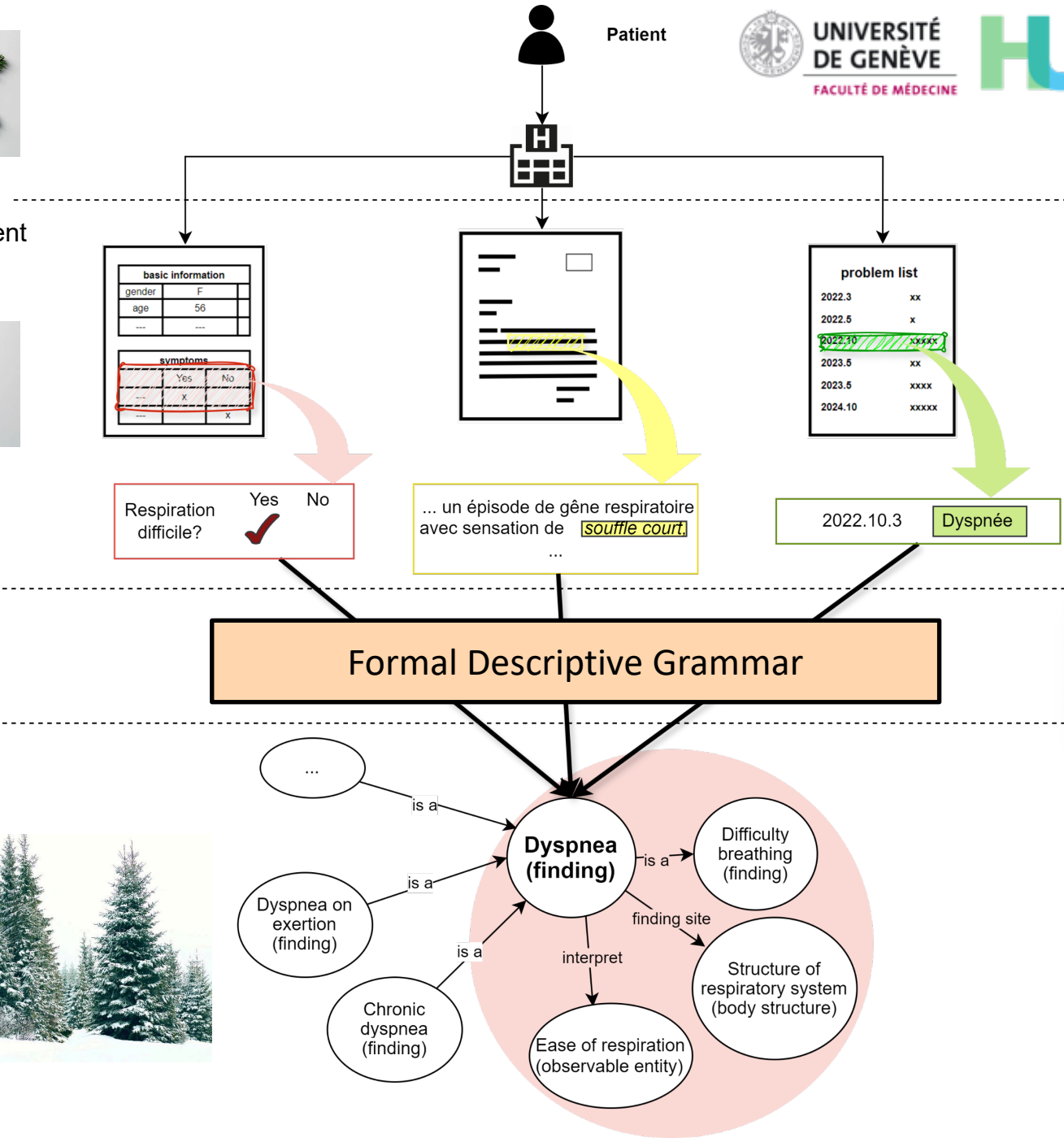
> 50% lossless dimensionality reduction

Usage of **post-coordination (compositional grammar)** when exact concept mapping is not available e.g with novel virus variants



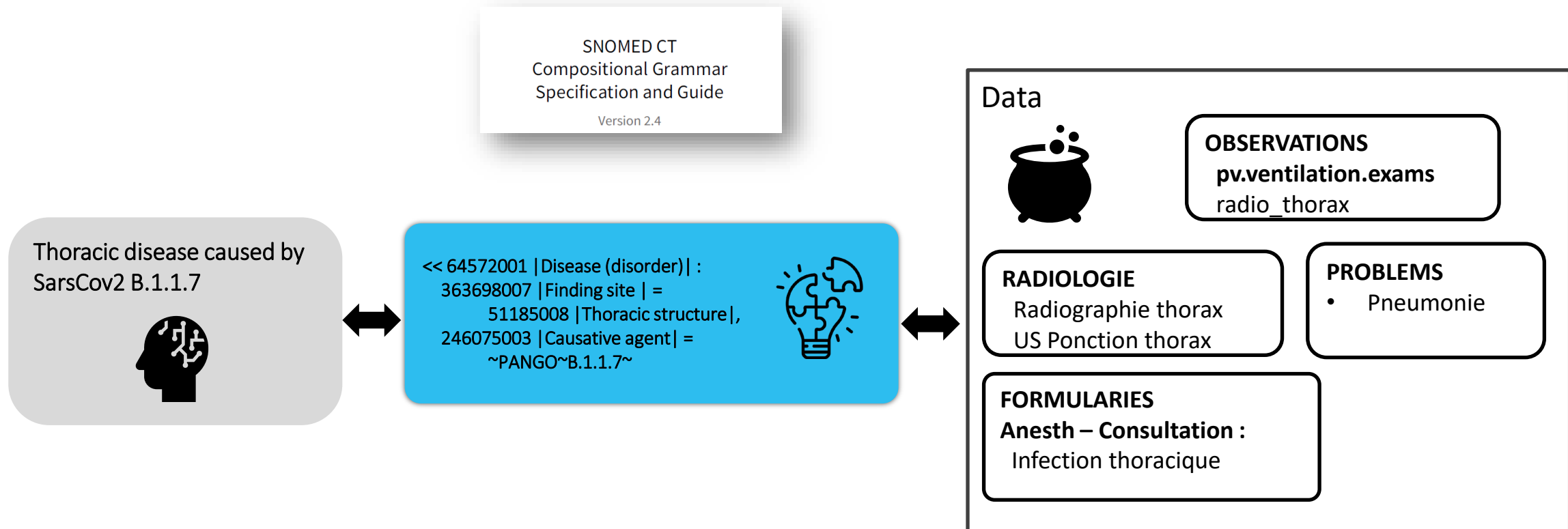
Data augmentation

- Link to **more than 350k concepts**
With **more than 100 relation types**
- ✓ Easier transfer to other ontologies
 - ✓ Easier transfer to data models
 - ✓ Enabling semantic driven further dimensionality reduction (e.g. aggregating all data elements expressing any hand bones)



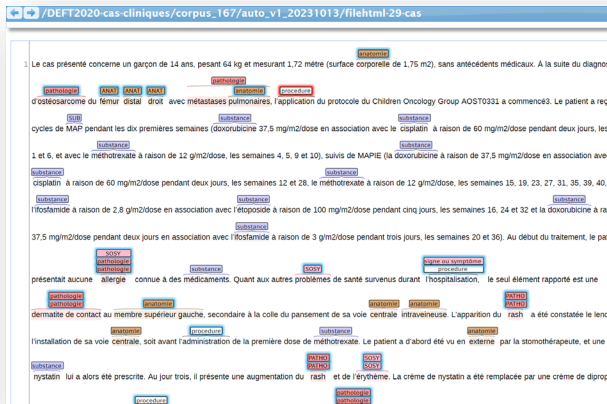
SNOMED CT
Compositional Grammar
Specification and Guide
Version 2.4

Passer des données à l'information



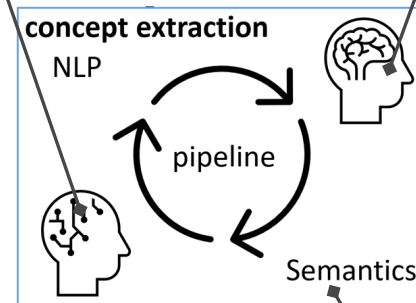
Automatic annotation (labelling)

Human evaluation



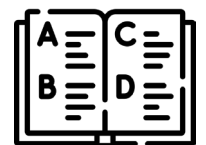
Manual annotation

Medical Experts



Building resources

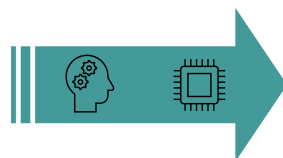
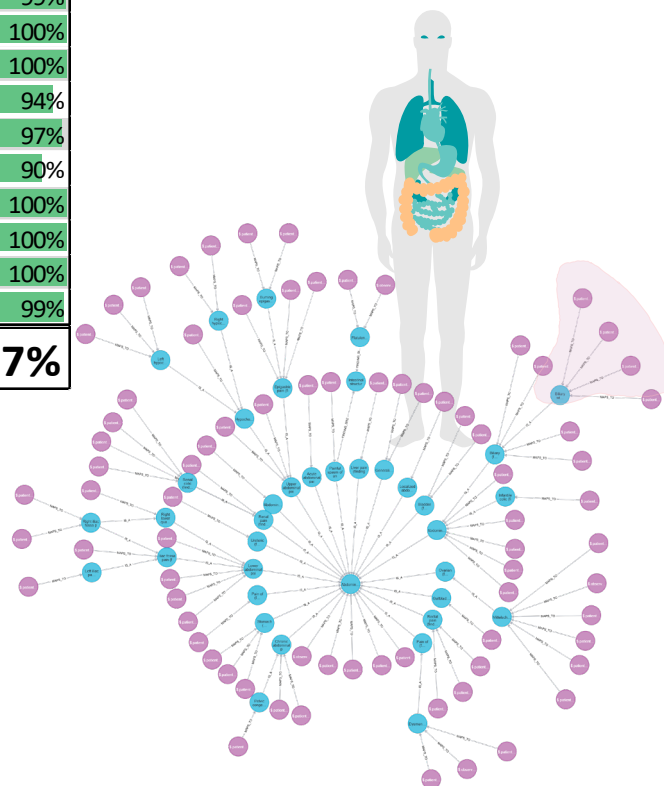
Reuse of manual work



Source	# ID	# coded	%coded	%inst_coded
Formulaires	178 584	34 332	19%	62%
Prescriptions	47 974	9 837	21%	98%
Act	46 806	5 777	12%	78%
Tasks	45 278	1 017	2%	91%
Problems	26 086	25 233	97%	99%
Clinisoft	15 527	6 841	44%	95%
Laboratoires	14 214	7 180	51%	95%
Metavision	6 956	2 137	31%	47%
Opera	5 307	760	14%	91%
Patient values	2 368	739	31%	93%
Radiologie	2 176	1 960	90%	99%
Aria	915	915	100%	100%
Équipements	813	699	86%	100%
Cinesio	810	801	99%	94%
Épisodes de soins	492	386	78%	97%
Échelles cliniques	454	220	48%	90%
Patient	366	359	98%	100%
Transmissions ciblées	264	264	100%	100%
Vaccinations	210	210	100%	100%
Triage	37	36	97%	99%
Total	395 637	99 703	25%	87%

Categorical
data

narratifs

Directed Acyclic
Graph (DAG)

1.7 million patients

12 billion instances

Patient	Problème	Date
Gaudet C.	Lumbago	18.11.2024

99'000 distinct data elements

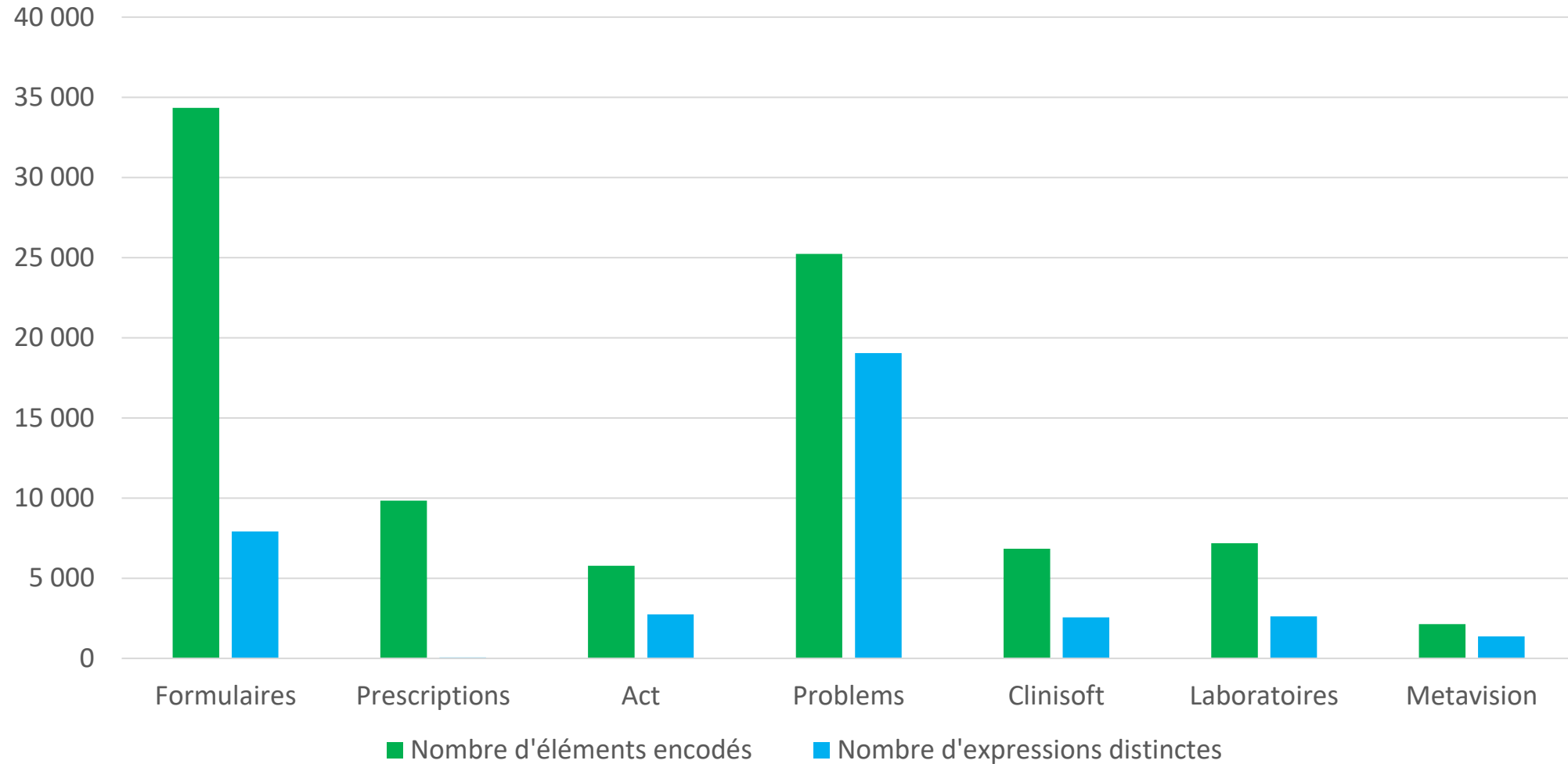
Type	ID	Label
Problème	22177	Lumbago

40'000 distinct expressions

279039007 | Low back pain (finding)|

Connected in a semantic/knowledge graph

22253000 | Pain (finding)|



- 60'000 expressions
- Built from how clinicians are reporting
- Mappings
 - SNOMED CT, ICD-10, ATC, CHOP, ICPC2, internal classifications



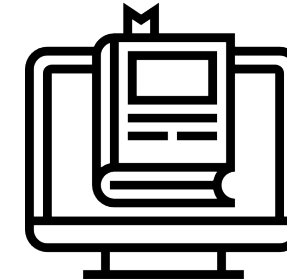
Médecine aiguë



Bloc opératoire



Réhabilitation



RECIT



Nutrition



Urgences



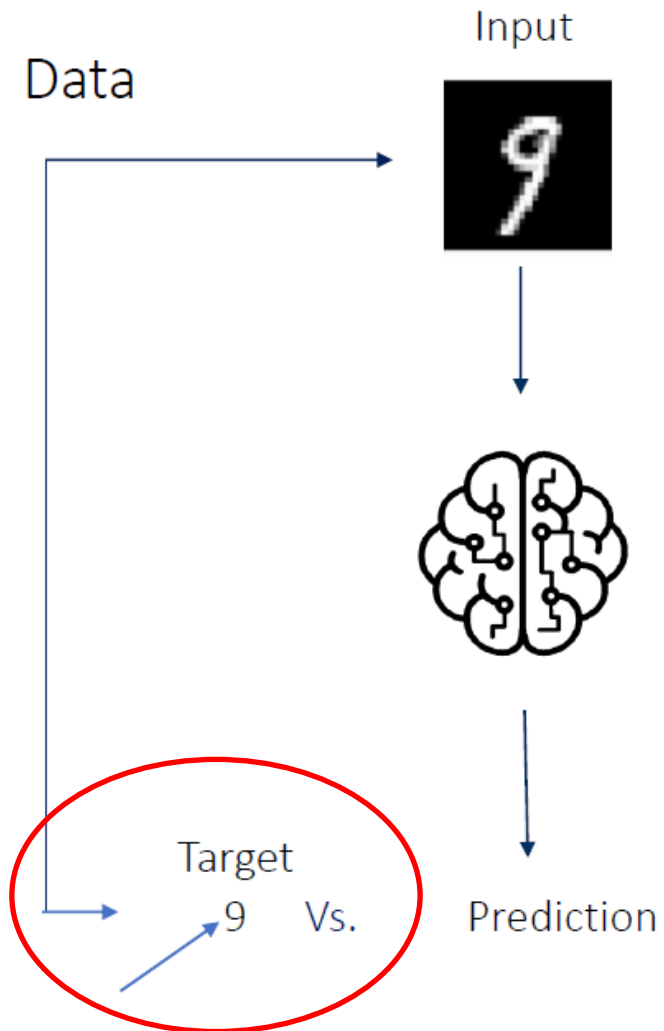
Antibiovigilance

PROBLEM ORIENTED PATIENT REPORTING

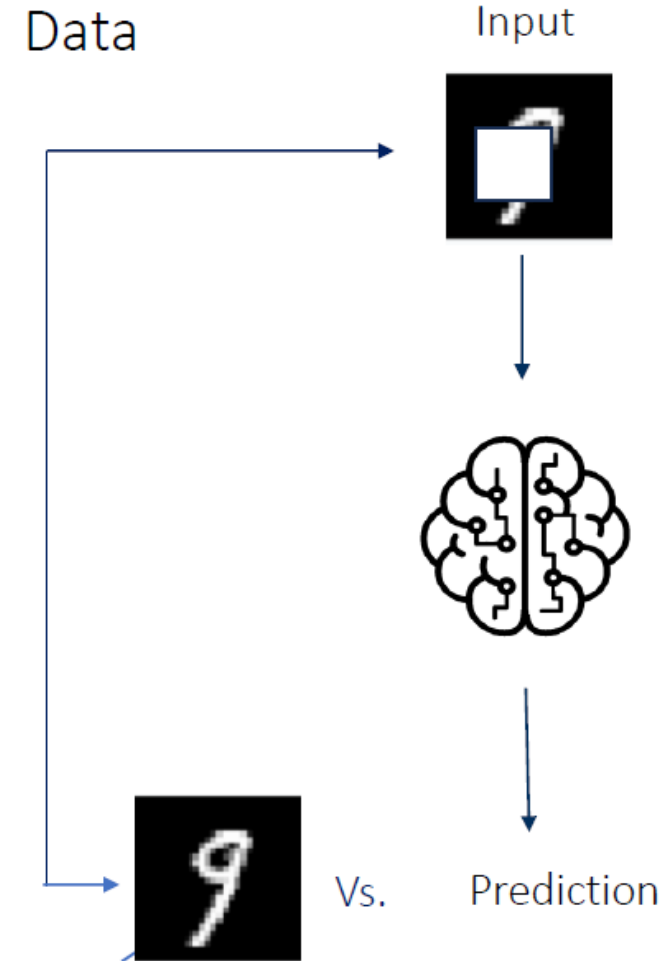
Libellé	Libellé
Inf	avc
HUG (300)	HUG (43)
Infarctus du myocarde	Accident vasculaire cérébral ischémique sylvien
Infection urinaire aiguë	Suspicion AVC fosse postérieure
Infarctus cardiaque	Suspicion d'AVC ischémique sylvien gauche
Infection à SARS-CoV-2 (COVID19)	Accident cérébral
Infection	Attaque cérébrale
Influenza	



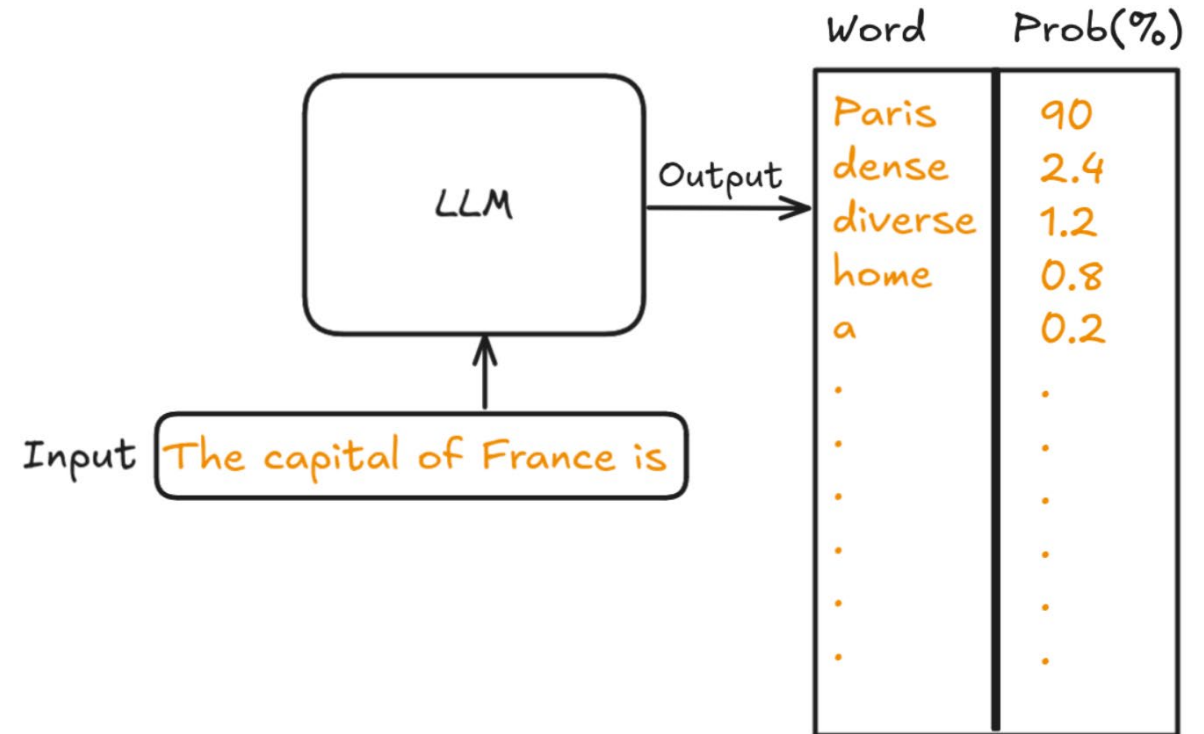
«self-supervised learning» revolution (SSL)

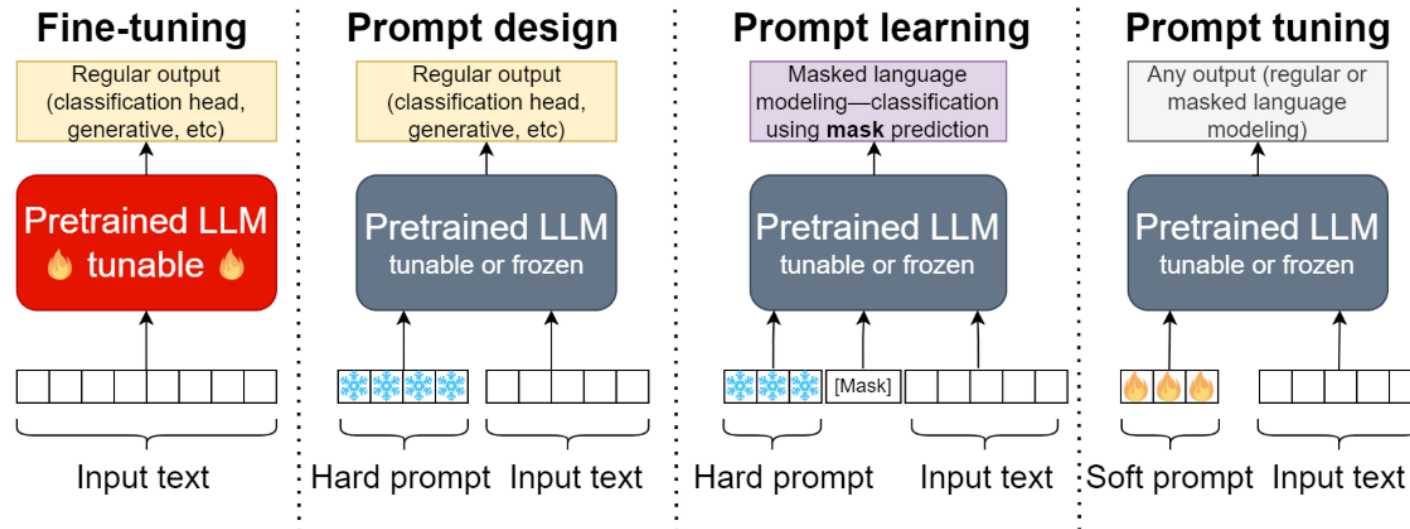


Annotation manuel



Language models





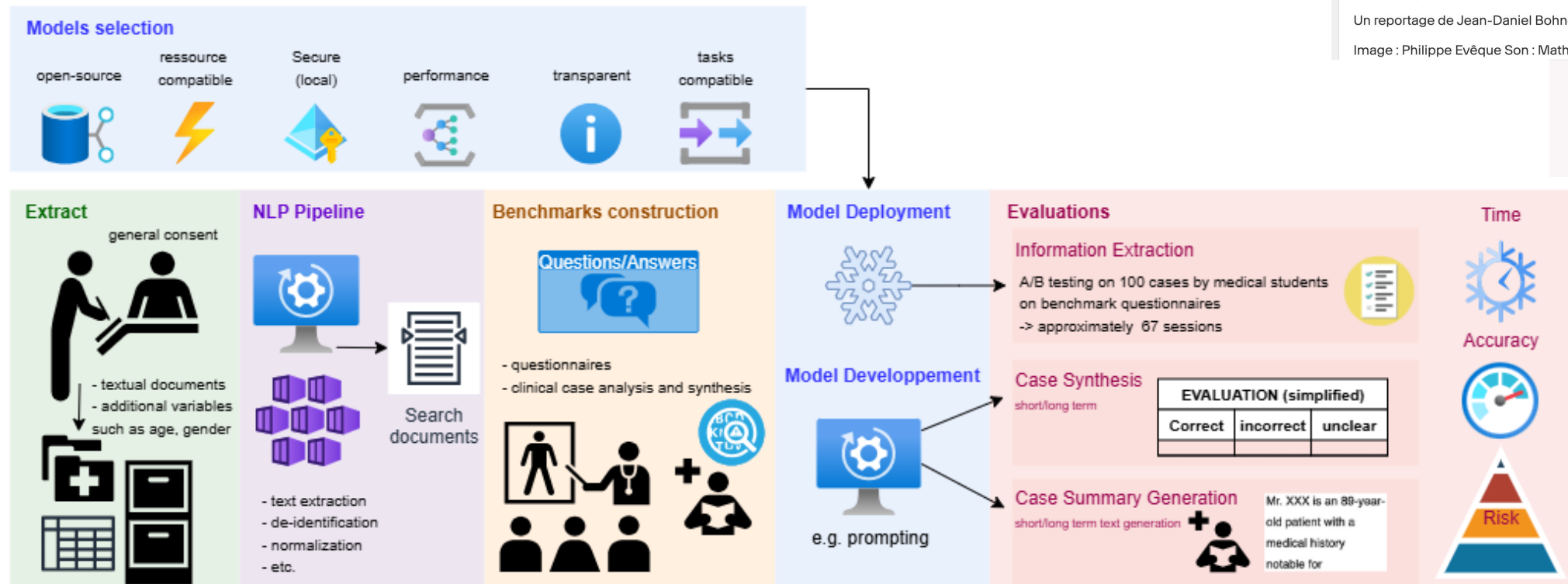
ChatGPT 4o >

draw one human female hand with only 3 fingers



Here is the image of a human female hand with only three fingers, depicted in a natural and lifelike pose. Let me know if you'd like further modifications!

Methodically assess the impact of AI's new hopes for clinical documentation

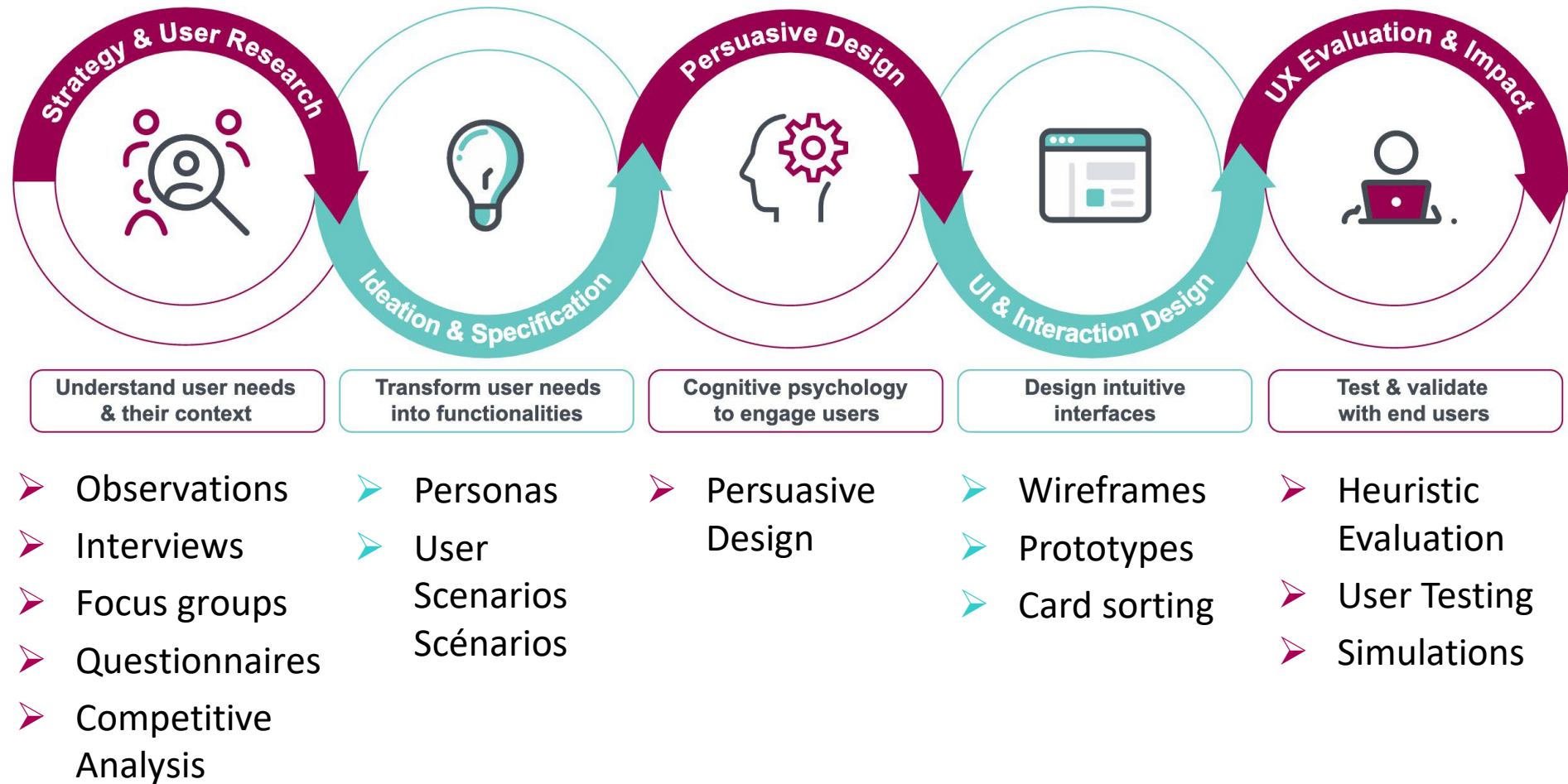


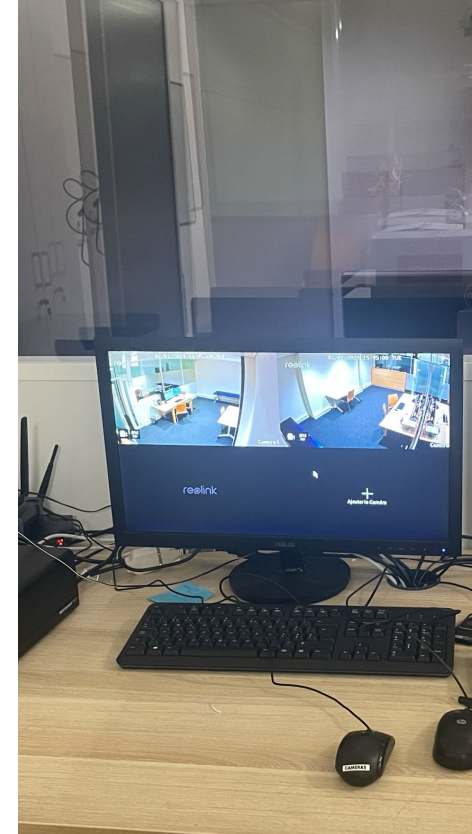
Devenir médecin ne fait plus vraiment rêver. Qu'ils soient praticiens, en formation ou étudiants, les médecins sont toujours plus nombreux à se détourner de la profession. Trop d'heures devant l'écran, trop peu avec les patients, des horaires interminables, la crainte de l'erreur médicale. Autant d'éléments mettant en tension un système médical qui risque de devoir affronter un sérieux problème de relève.

Générique

Un reportage de Jean-Daniel Bohnenblust

Image : Philippe Evêque Son : Mathilda Angullo Montage : Alice Naylor Illustration

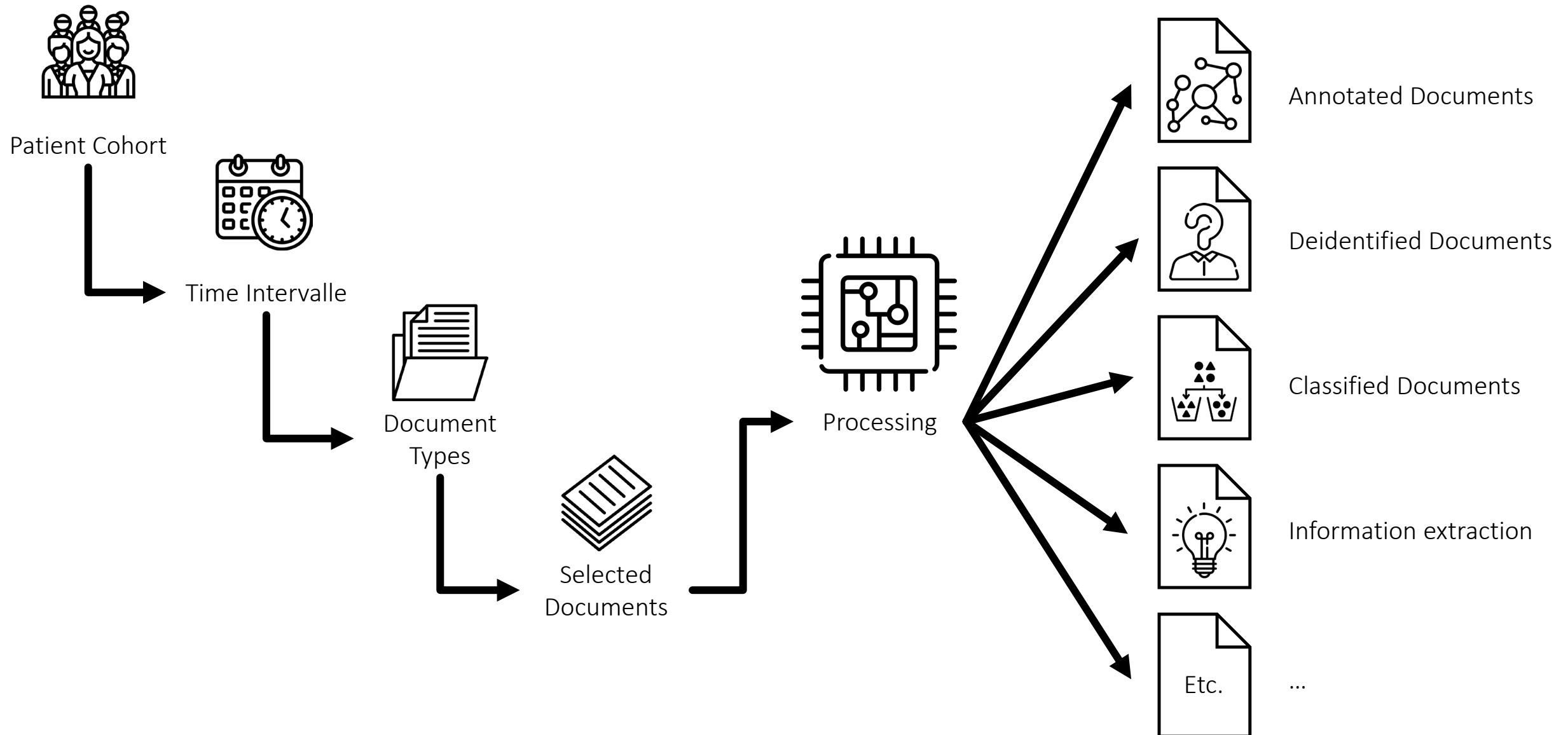




Cameras



Eye tracking



The SIMED Natural Language Processing Pipeline



The SIMED NLP Pipeline covers all medical specialties, from acute care to rehabilitation.

1.25 billion words
with clinical pertinence

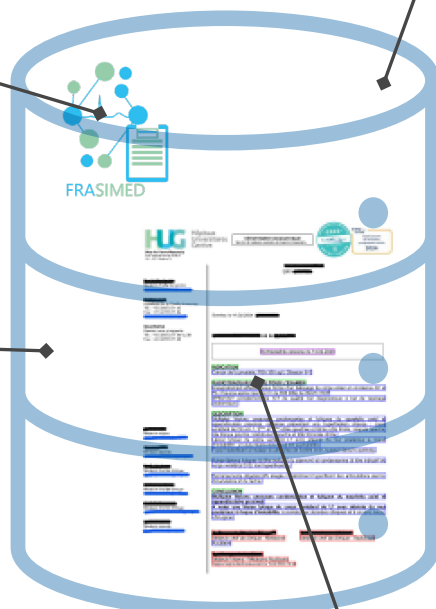
80 million documents
free text in medical French,
layout with affiliations, tables,
images



2.17 million documents
in high quality machine-readable
format

FRASIMED
Automatic translation
& annotations of
external datasets

MedFrenchMark
Benchmark with
relevant clinical
questions for testing
AI models



UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



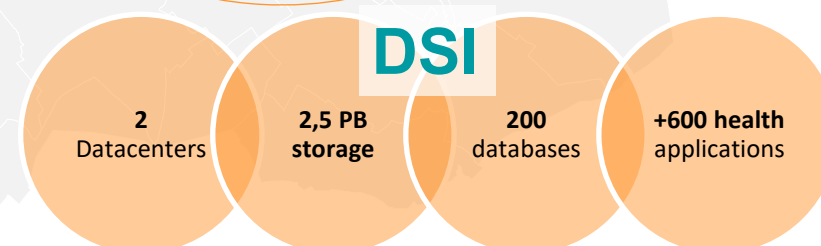
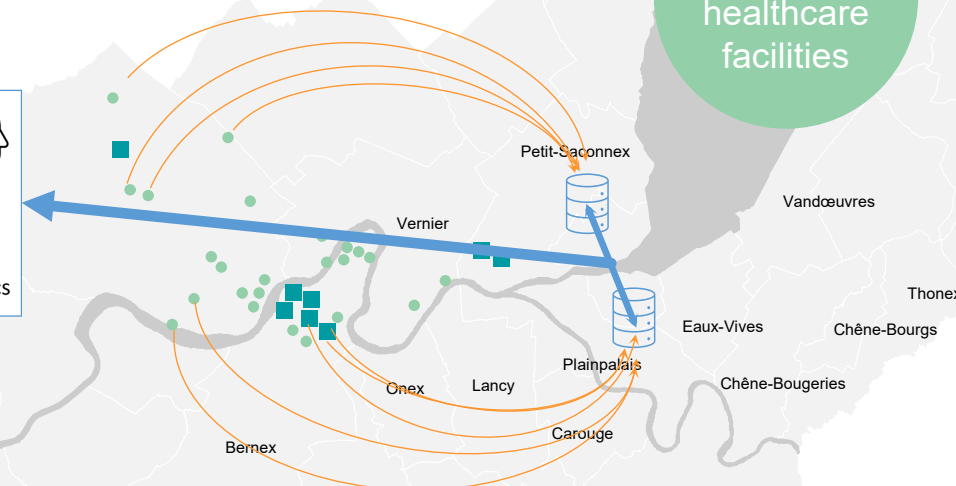
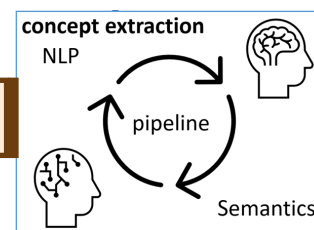
Hôpital
Universitaire
Genève

1.7 million patients
In the HUG Network
databases



■ Hospital sites
● Healthcare facilities

+20
healthcare
facilities



Tools for research and deployment



The SIMED pipelines are designed to enable research and fast, secure and reproducible integration to respond to the fast AI fast evolving fields.

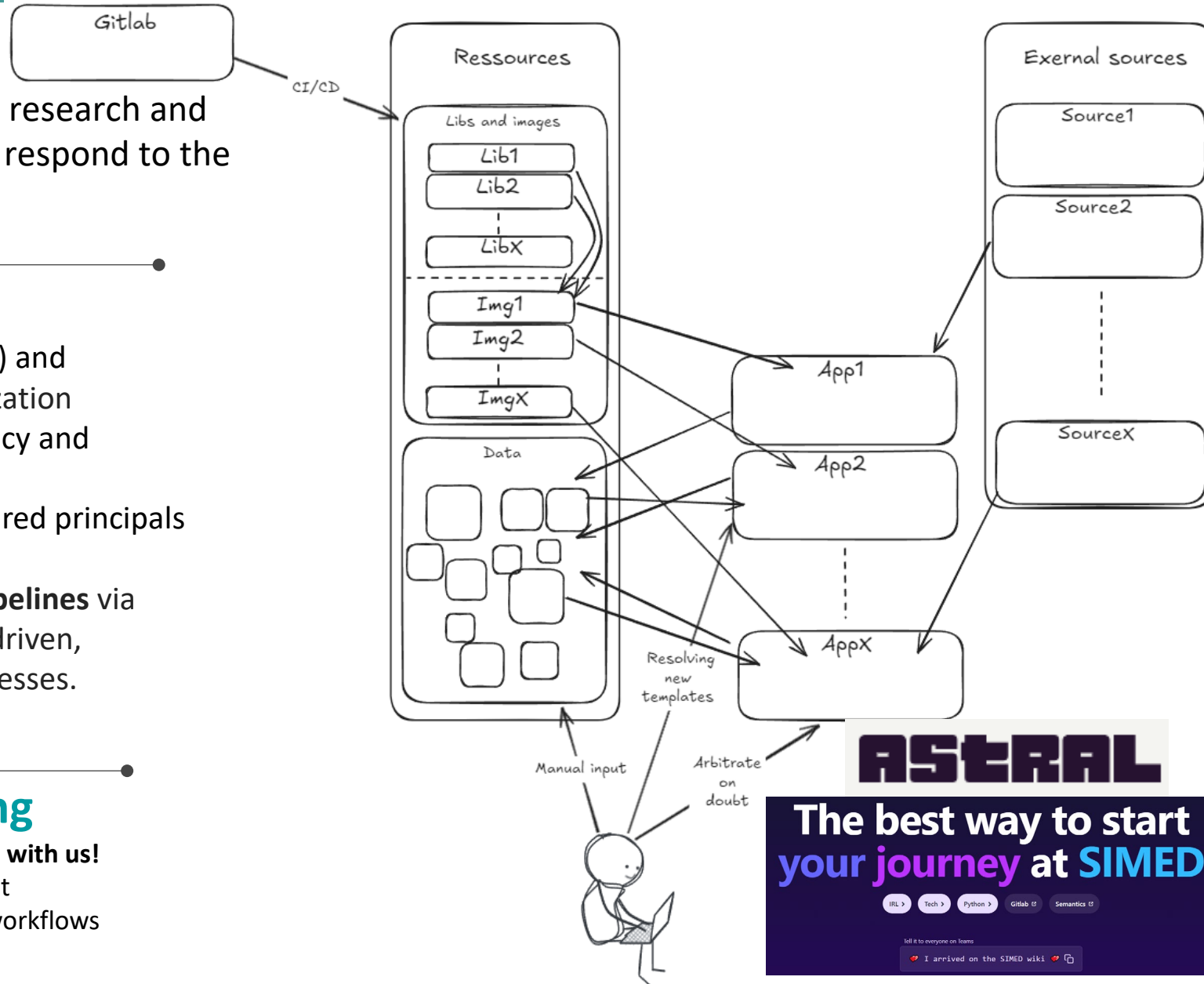
System Principles

- ✓ **Code reuse** via continuous integration (CI) and continuous delivery (CD), and containerization
- ✓ **Version Consistency** via formal dependency and environment definition
- ✓ **Perfect traceability** using blockchain inspired principals
- ✓ **Balance** of Freedom and standardization
- ✓ **Transparent, traceable, and auditable pipelines** via blockchain industry principles e.g. event driven, immutable, cryptographically signed processes.

SIMED streamlined onboarding

It's a lot to take, we know, let's help you get on board with us!

- ✓ Few hours rather than weeks to setup environment
- ✓ Internal wiki to quickly become familiar with our workflows
- ✓ Best practice common updates across projects



Enabled Projets

HUG
Direction médicale

LPH
Liste des problèmes HUG

HUG
Service d'oncologie de précision

SPO
Extraction automatique et classification de documents

HUG
Médecine nucléaire et imagerie moléculaire

NLP Pipeline
Extraction automatique et classification de documents

CHUV
Service de médecine interne

LUCID
Détection des Low Value Care

HUG
Service des urgences

Euro DEN Plus
Détection de consommation de drogue

Bern Inselspital

INFRA
Détection automatique du sepsis

HUG
HERO
Automatic cohort builder

HUG
XpliCAI
Comprendre la transparence



HUG
SMPR

STIGMA
Impact du langage stigmatisant

HUG
Service de radiologie

SUBREAM
Extraction automatique et classification de documents

UNIVERSITÉ DE GENÈVE
UNIGE
SDC

gdhub
Geneva Digital Health Hub
Building a Science-Informed, Equitable Future for Digital Health

AAL
PROGRAMME

engAGE Guardian H2HCare
Evaluation de systèmes robotisés pour les personnes âgées

CHUV
Biomedical Data Science Center

S4
Smart SNOMED Search for SPHN

HUG
Service de néphrologie et hypertension

AIAKI
Prédiction de l'insuffisance rénale aiguë

SHIELD
IA et données multimodales : prévention personnalisée des maladies cardiovasculaires et du diabète

HUG
CliniDoc Assistant
IA pour la navigation dans le dossier patient

HUG
DREAM
IA pour résumer le dossier patient

Enabled data reuse



UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



Hôpital
Universitaire
Genève

 **SIMED**  Knowledge graph
(40k dimensions, 12 billions instances)

 **SIMED**  NLP Pipeline
(2.17 millions documents
1.25 billion words)

 **SIMED**  Lexico-semantic ressources
(eg. morpho-semantems)

 **SIMED**  ECG database
(500k signals, 23k echocardios,
2k cardiac MRIs, etc.)

1.7 million patients

In the HUG Network databases

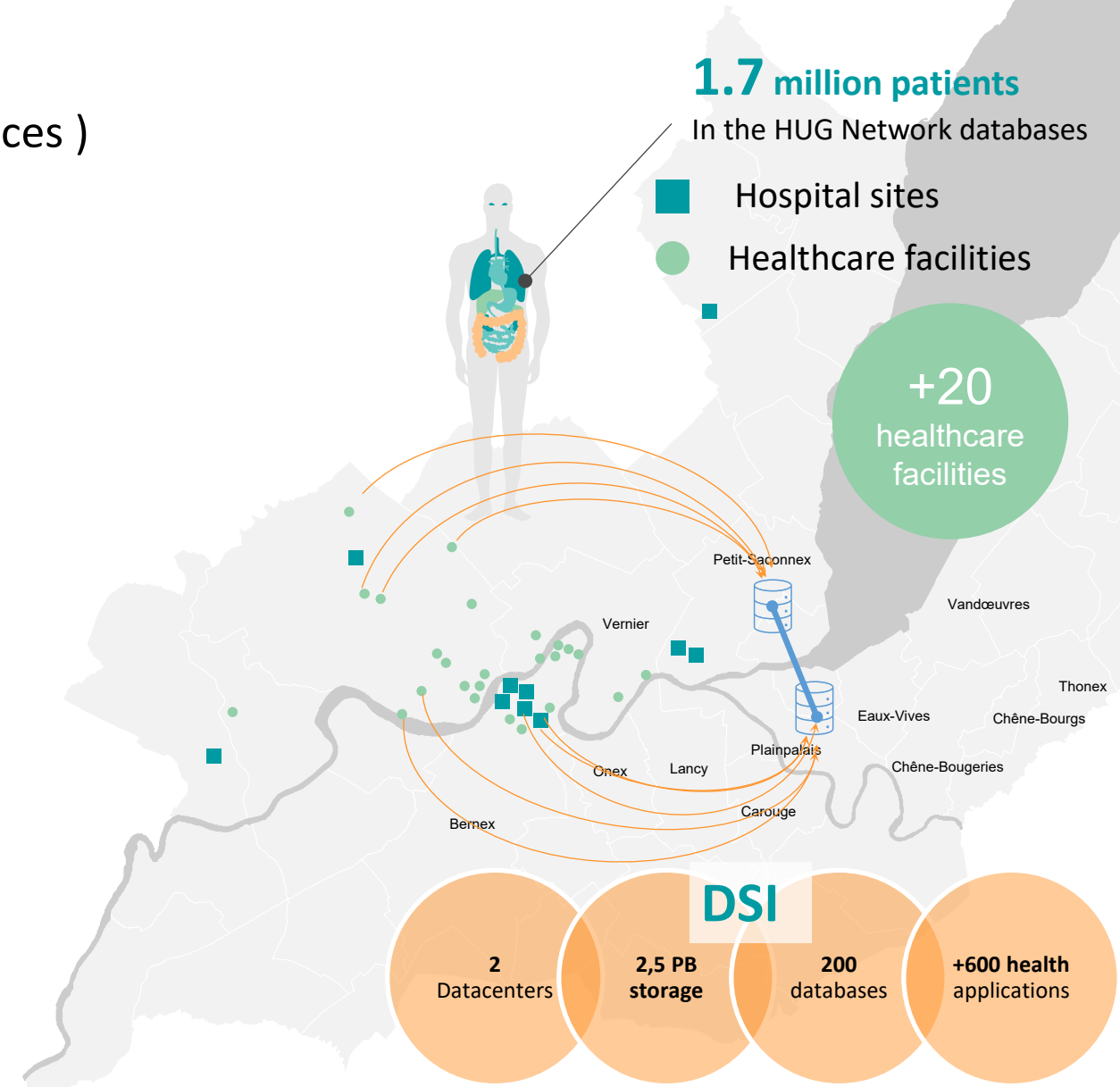


Hospital sites



Healthcare facilities

+20
healthcare
facilities



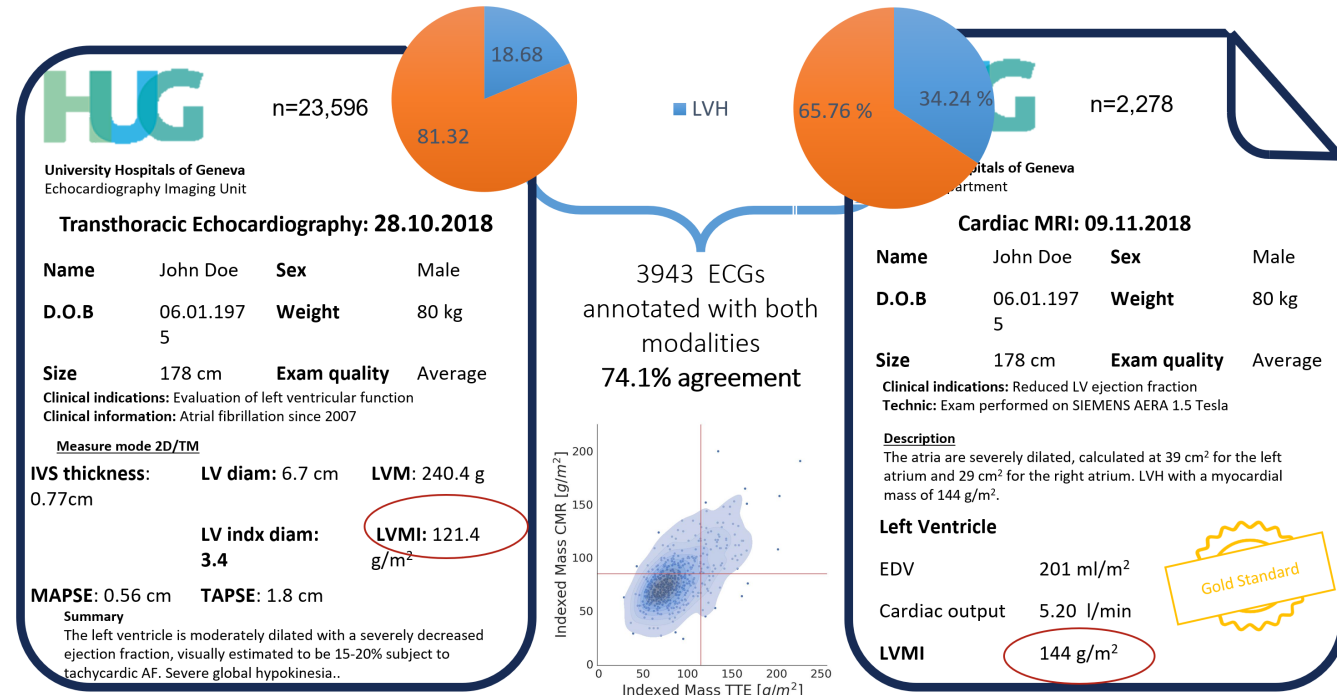
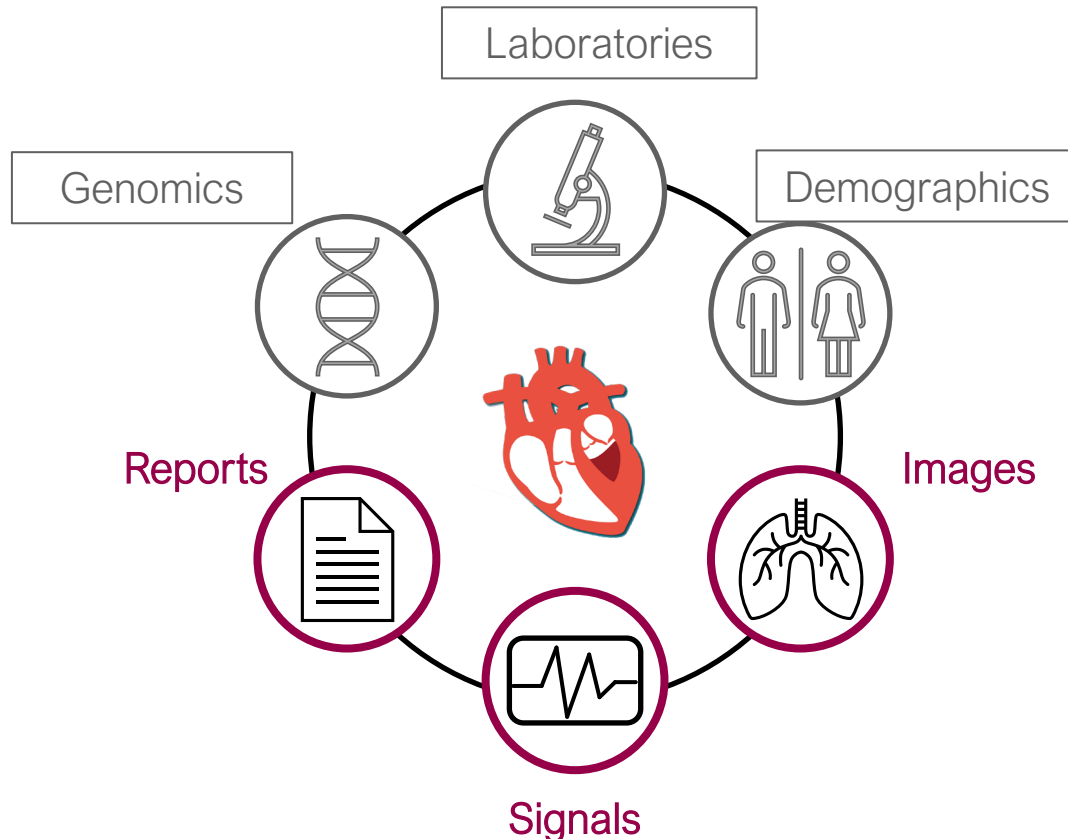
Multimodal and contextualised diagnosis



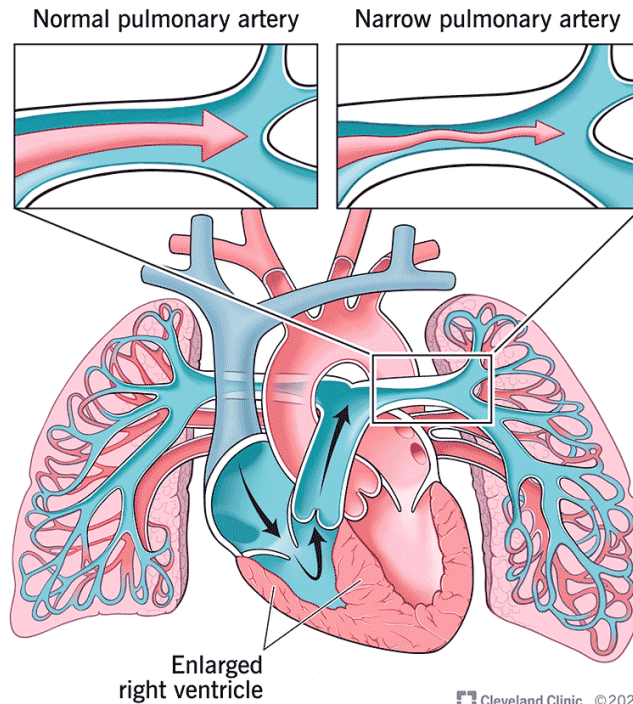
UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



Hôpitaux
Universitaires
Genève



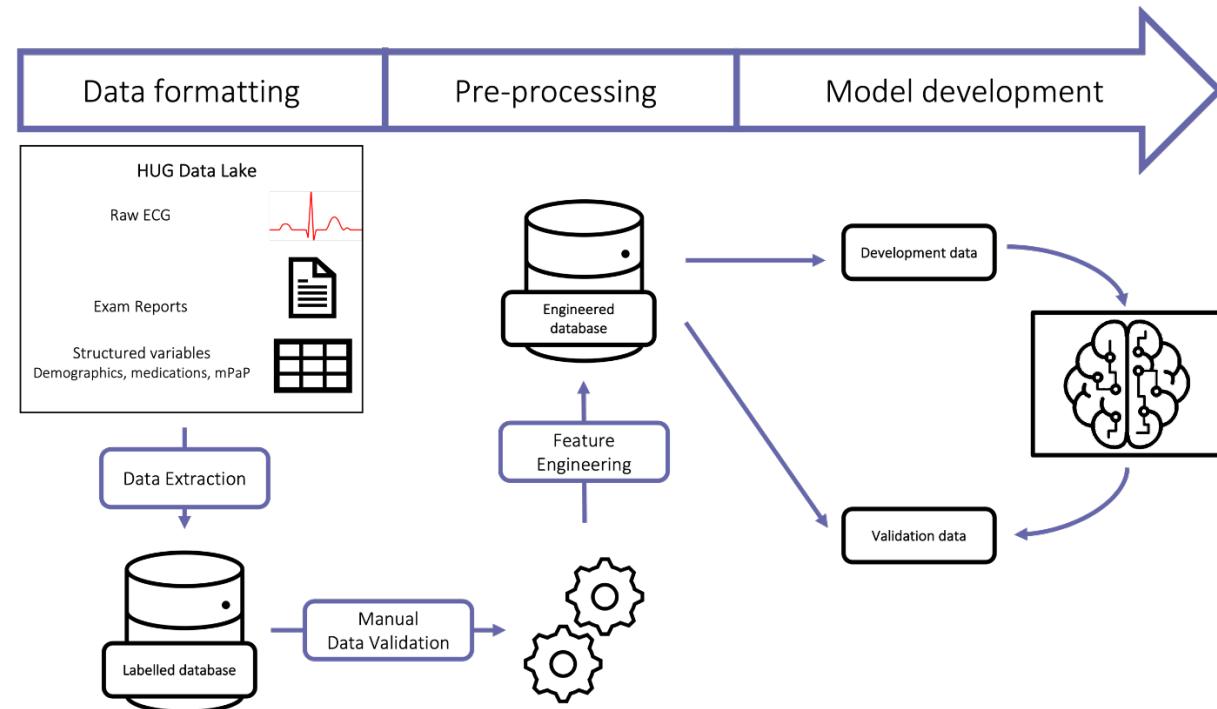
H. Turbé, M. Bjelogrić, M. Namdar M, Christian Lovis. Comparison of Echocardiographic and Cardiac Magnetic Resonance imaging for left ventricular hypertrophy diagnosis. J Am Coll Cardiol. 2023 Mar, 81 (8_Supplement) 1462. [https://doi.org/10.1016/S0735-1097\(23\)01906-X](https://doi.org/10.1016/S0735-1097(23)01906-X)



Cleveland Clinic ©2022



- 1% (10% above 65 year old)
- Invasive and late diagnosis
- Early diagnosis = more effective treatments



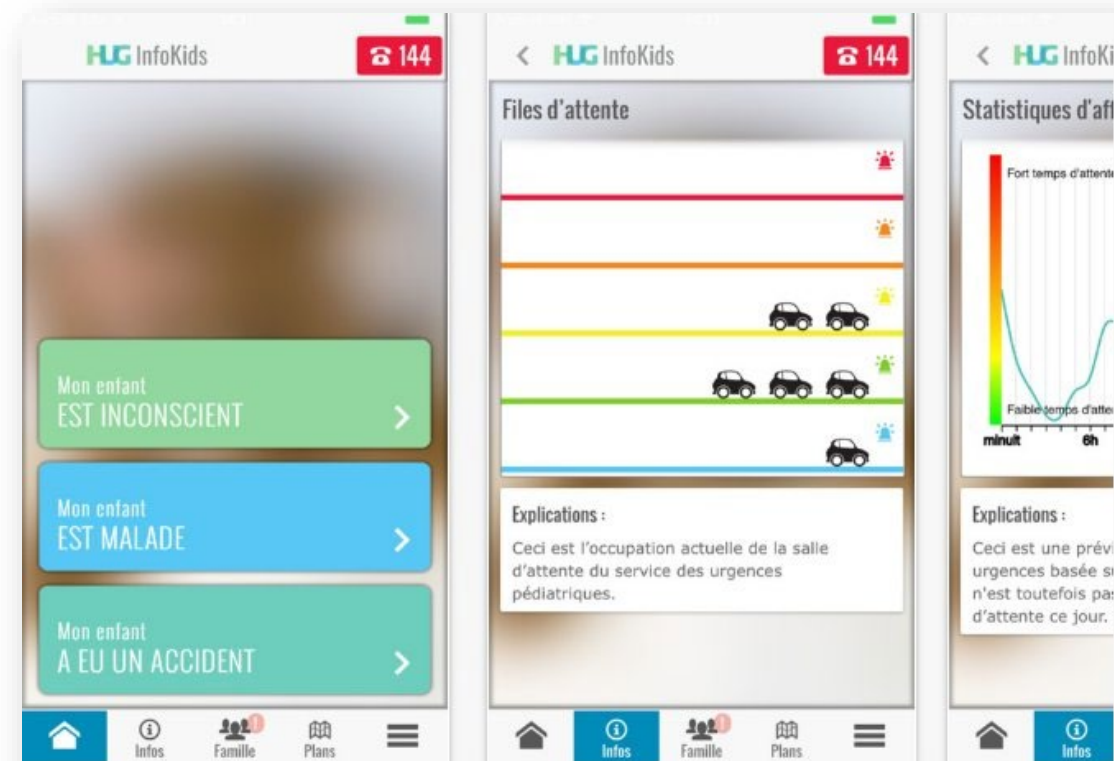
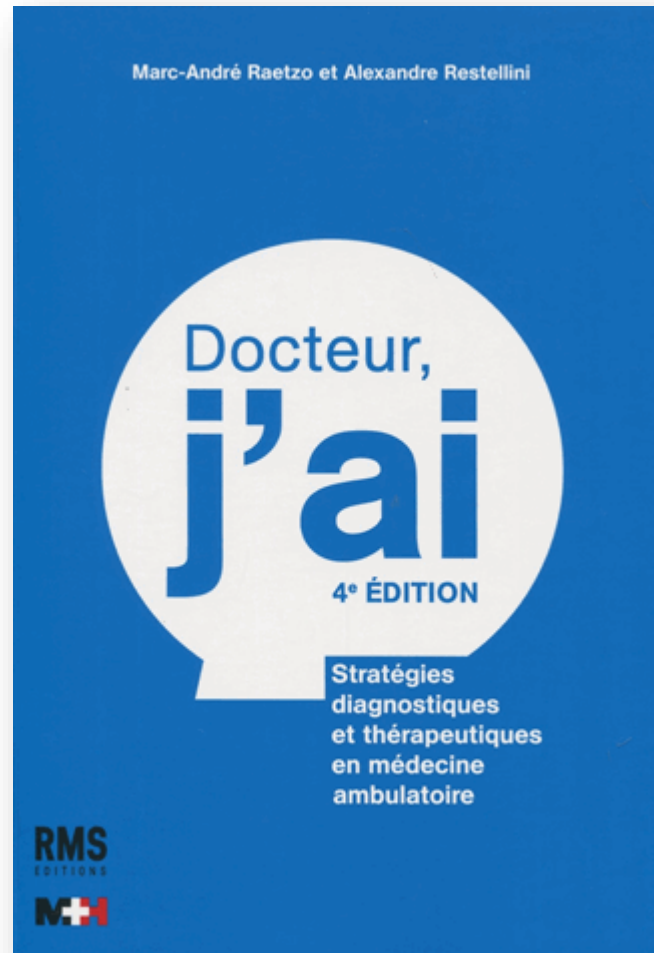


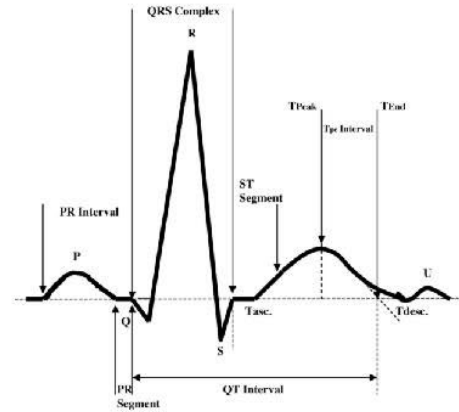
(Almost) real image of Stochastic Parrots generating outputs for large language model :) (generated with Stable Diffusion model).

The question we face is how to make sure the new AI tools are used for good rather than for ill. To do that, we first need to appreciate the true capabilities of these tools

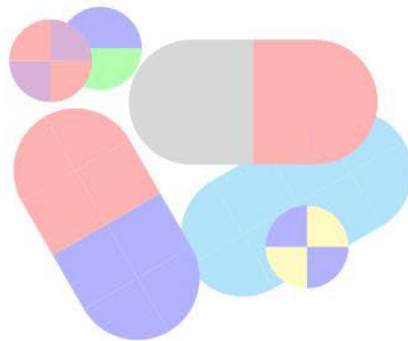
Yuval Noah Harari

Professor in the Department of History at the Hebrew University of Jerusalem
Author of Sapiens

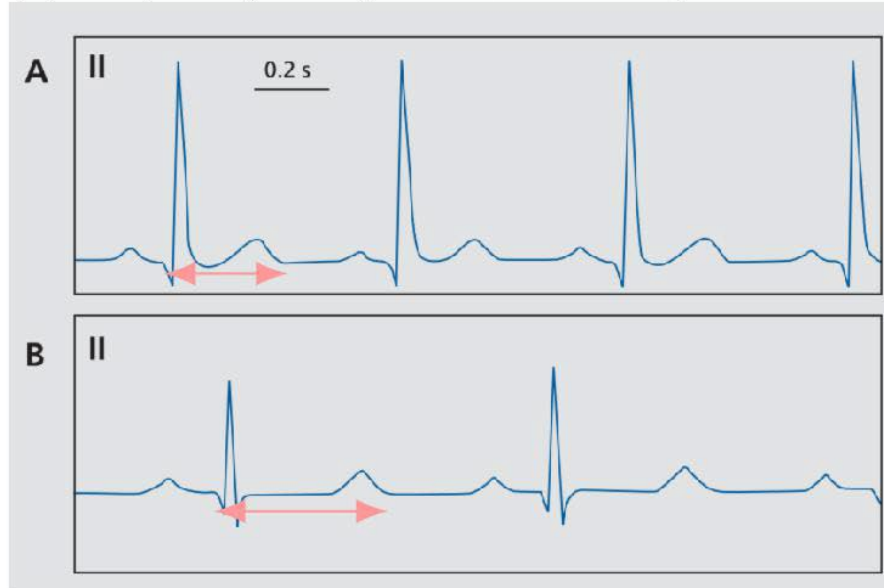




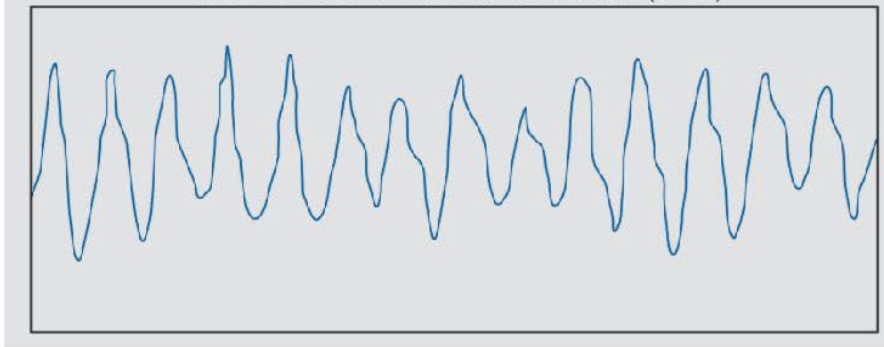
diLQTS

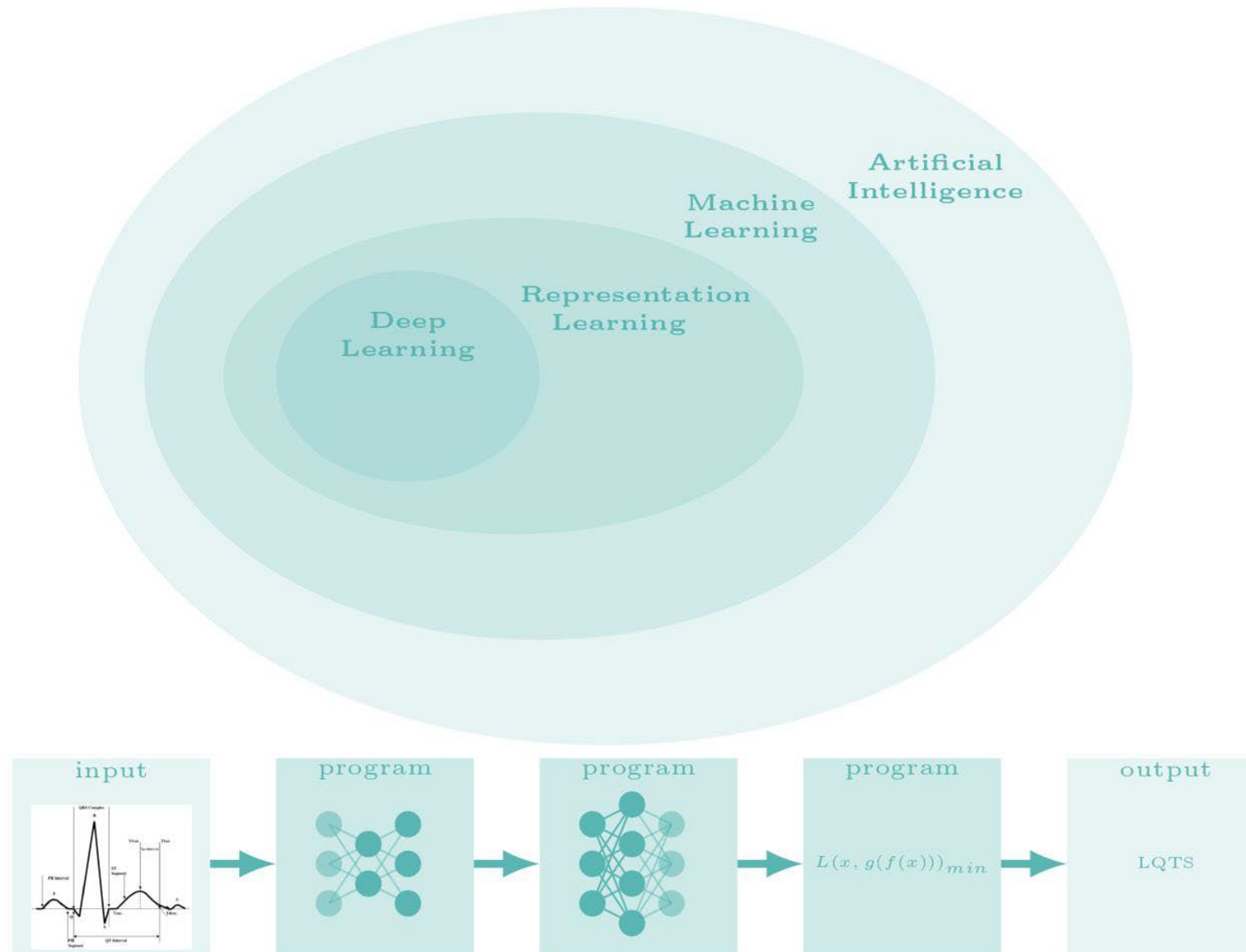


Girardin F, Sztajzel J. Cardiac adverse reactions associated with psychotropic drugs. *Dialogues Clin Neurosci* 2007;9:478-81.



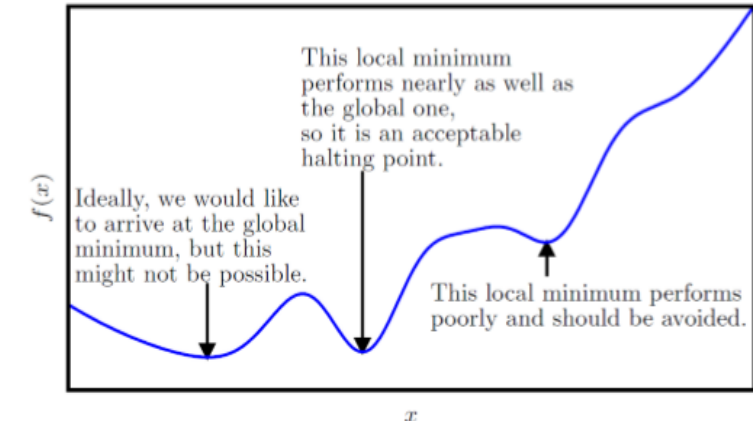
ECG lors d'une Torsade de Pointe (TdP)

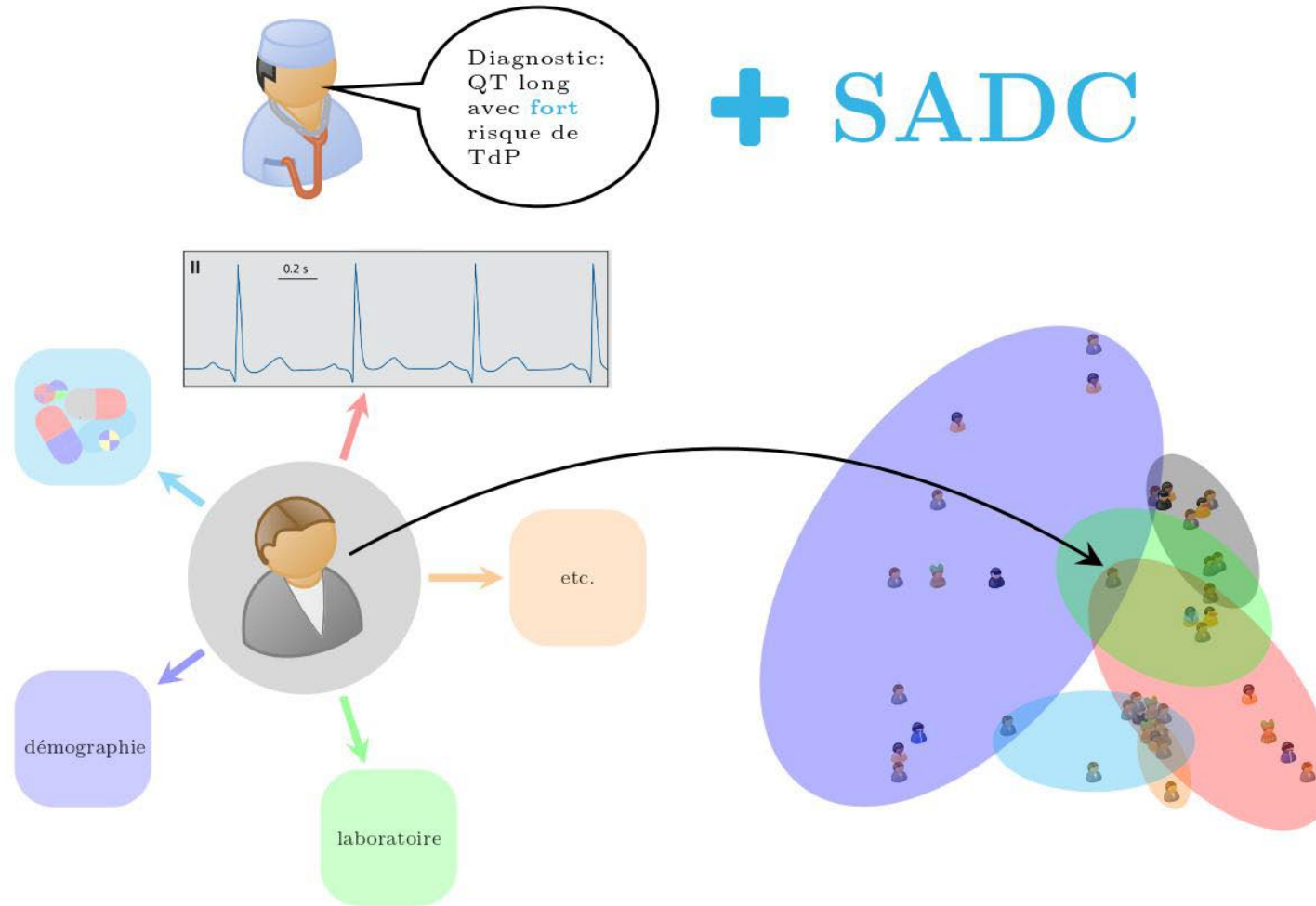


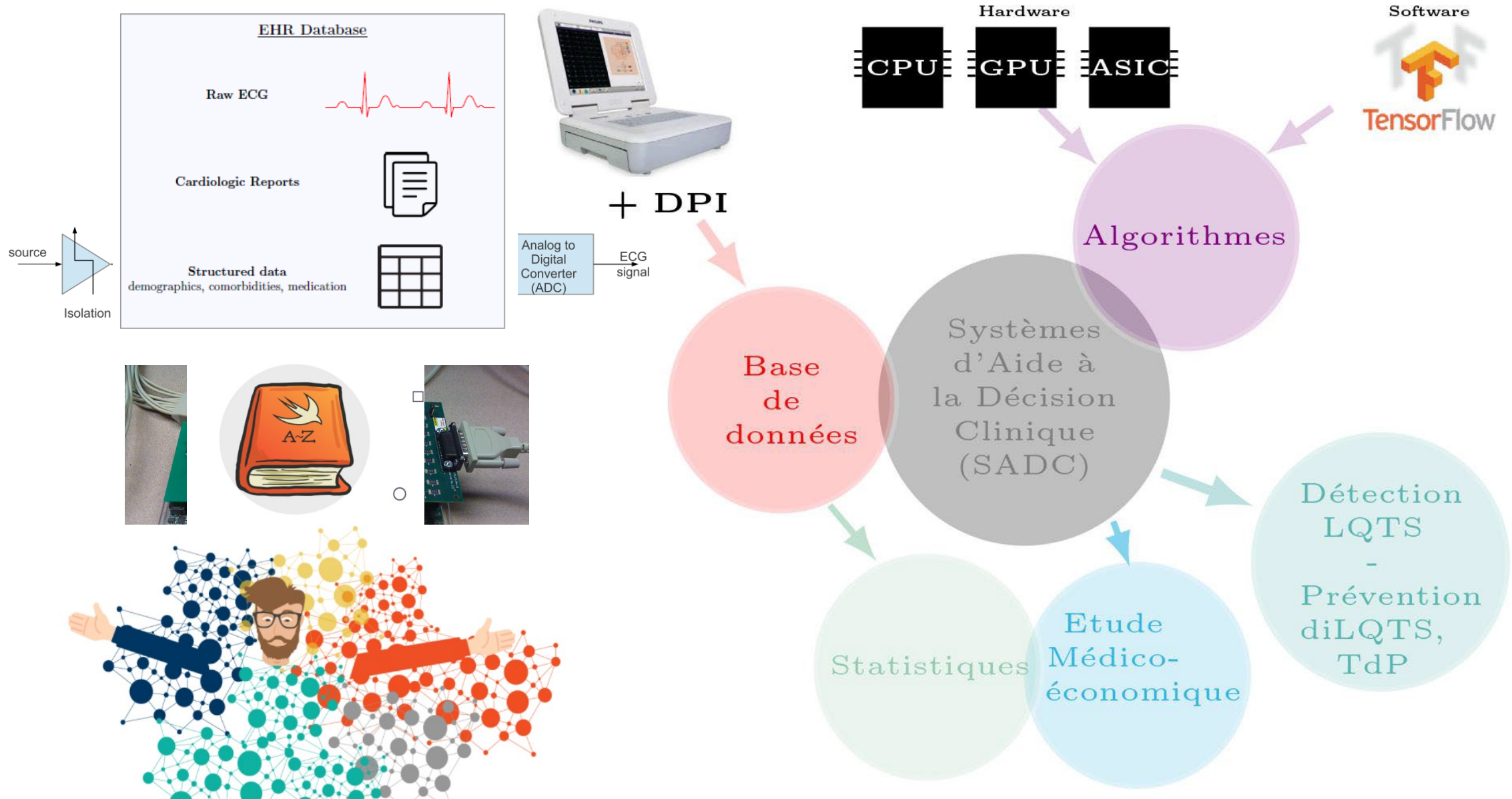


TRAIN	TEST
80%	20%

minimization (maximization)







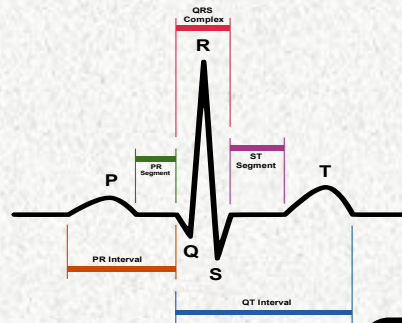
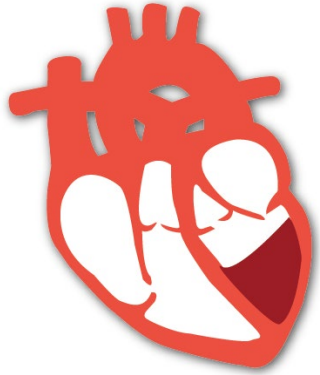
What do you choose



UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



Hôpitaux
Universitaires
Genève



Sokolow-Lyon:

$$S_{V1} + (R_{V5} \text{ ou } R_{V6}) > 35$$

Data

Rules /
Equations

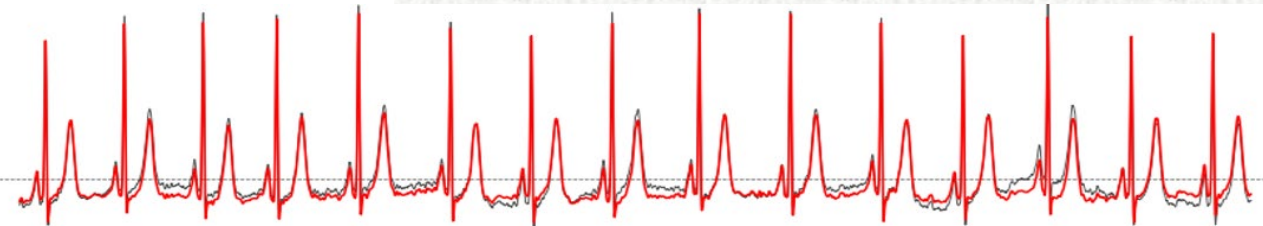
Results

TABLE 3

Diagnostic statistics of ECG-LVH criteria for Echo-LVH

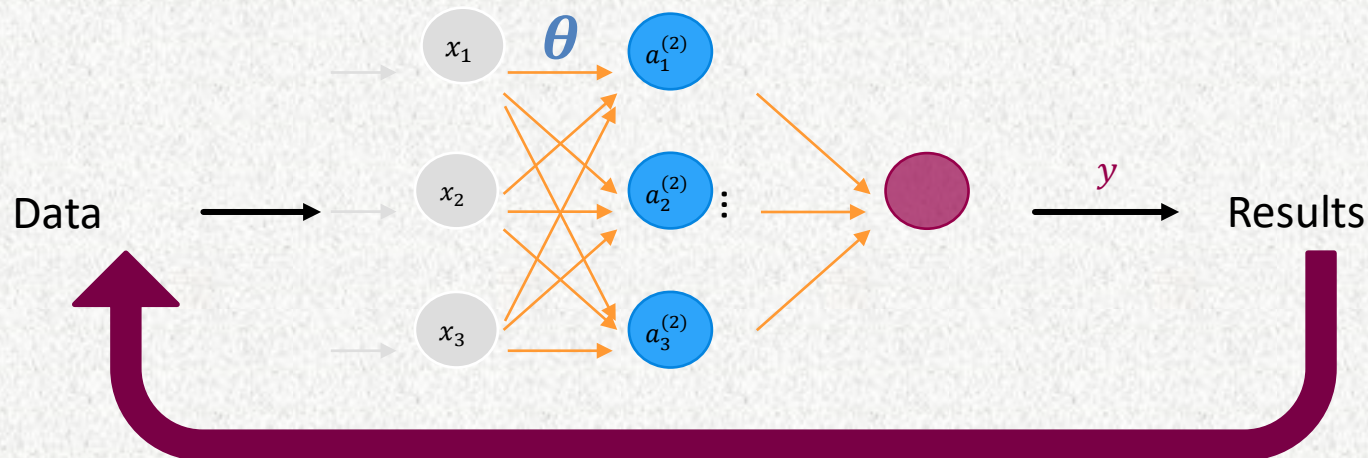
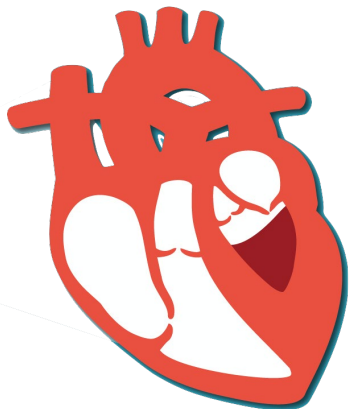
ECG criteria	Sensitivity(95%CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy(95%CI)
Pegueró-Lo Presti	29(26, 32)*	73(72, 74)*	12(11, 14)	89(88, 90)	68(67, 69)
Cornell voltage	12(9, 14)*	95(95, 96)*	24(20, 29)*	89(89, 90)	86(85, 87)
Cornell product	4(3, 5)*	96(96, 97)*	12(8, 16)	89(88, 90)	86(85, 87)
Sokolow-Lyon voltage	12(10, 15)*	89(89, 90)*	13(10, 15)	89(88, 90)	81(80, 82)
Sokolow-Lyon product	4(3, 6)	96(96, 97)*	13(9, 17)	89(88, 90)	86(85, 87)

100% explainable



LVH?

Which one do you choose?



0% explainable?

Classification Metrics

Accuracy 0.88

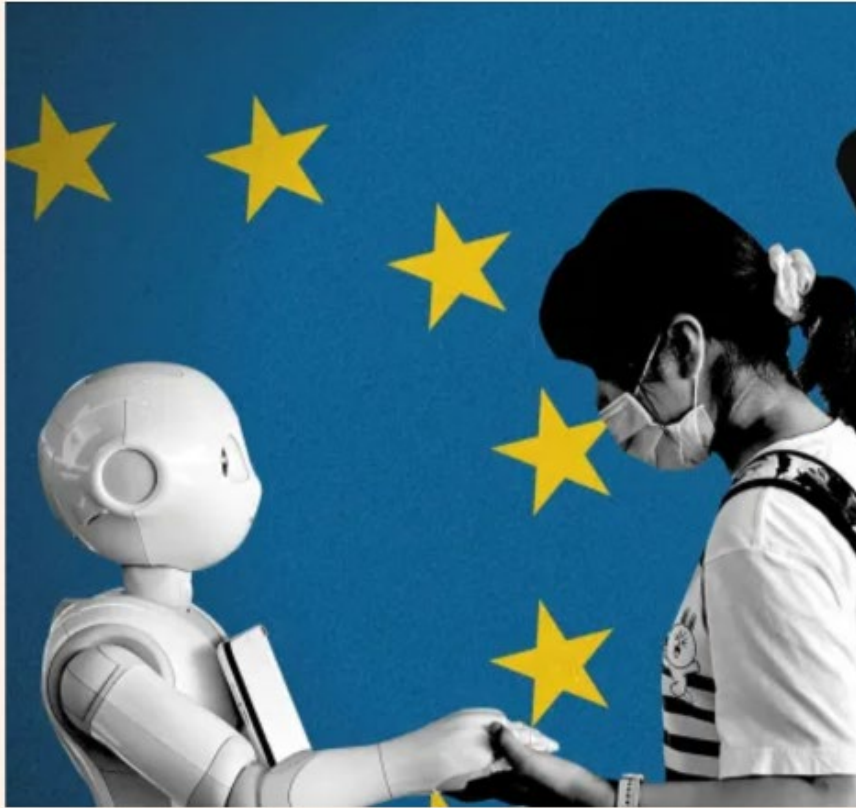
Sensitivity 0.80

Specificity 0.91

F1 0.83

Europe attempts to take leading role in regulating uses of AI

New rules would curb excesses while encouraging innovation with goal of a middle way between China and US policy

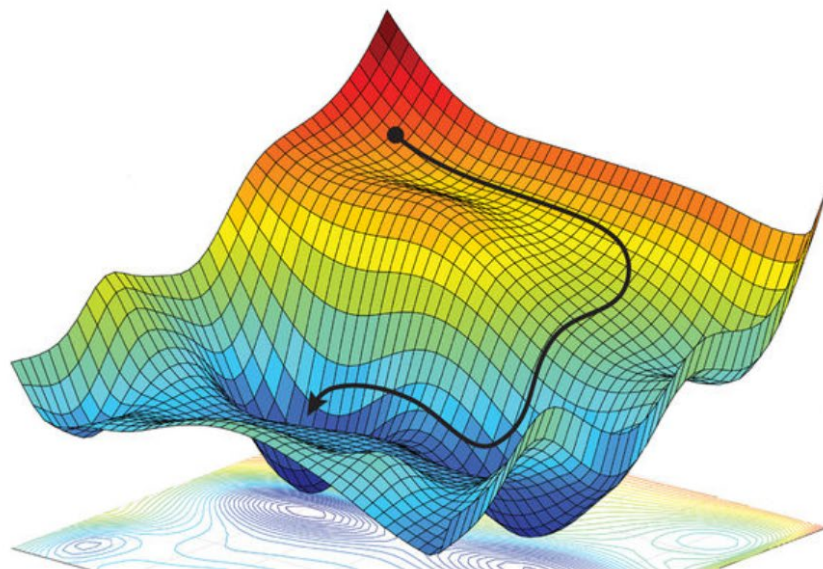


© FT montage; AFP/Getty

"critics say that, in trying to support commercial AI, the draft legislation does not go far enough in banning discriminatory applications of AI-like predictive policing, migration control at borders and the biometric categorisation of race, gender and sexuality. These are currently marked as **high-risk applications**, which means anyone deploying them will have to notify people on whom they are being used, and **provide transparency on how the algorithms made their decisions** — but their widespread adoption will still be allowed, particularly by private companies"

Explainability

SHARE



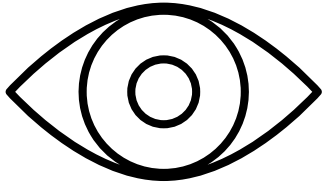
Gradient descent relies on trial and error to optimize an algorithm, aiming for minima in a 3D landscape.
ALEXANDER AMINI, DANIELA RUS. MASSACHUSETTS INSTITUTE OF TECHNOLOGY, ADAPTED BY M. ATAROD/SCIENCE

AI researchers allege that machine learning is alchemy

By Matthew Hutson | May. 3, 2018, 11:15 AM



AI Awareness



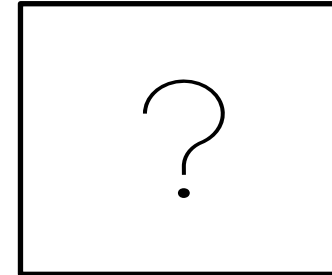
Is there AI in the product ?

AI Data Transparency



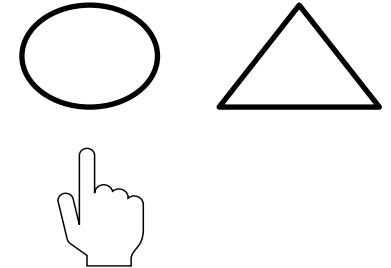
Which data are used in the product?

AI Model Explainability



How does the model work?

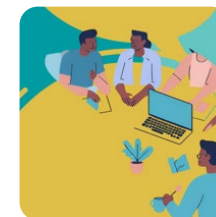
AI Outcome Explainability “Interpretability”



Why the model gave this answer?

XplicAI

Which AI transparency do we want at HUG?



Fondation
privée des

HUG

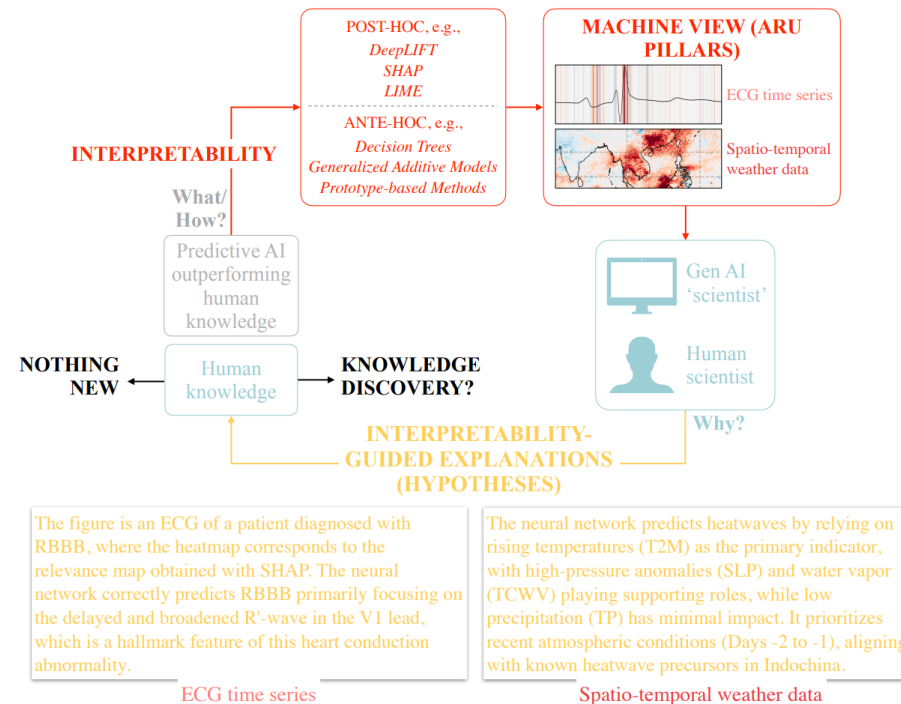
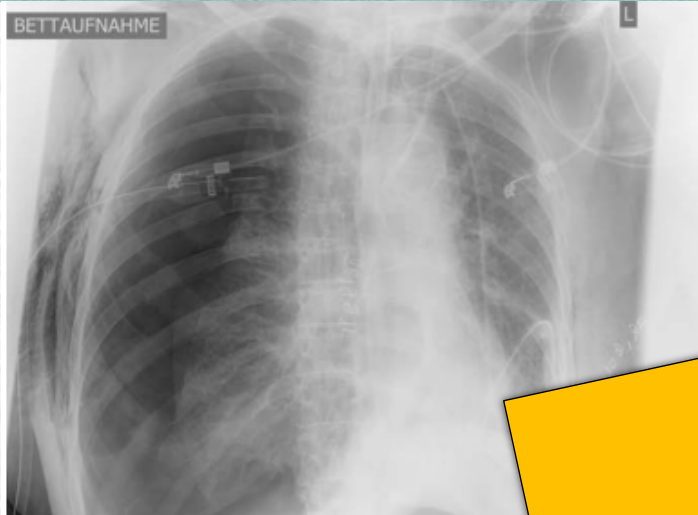


Figure 2: XAI for Science workflow. A predictive AI outperforms human expertise (grey box). Interpretability is then applied to generate the machine view (red boxes) that meet accuracy, reproducibility, and understandability (ARU) criteria. These interpretability results are then analyzed by human domain experts or generative AI models (green box) to derive interpretability-guided explanations (IGEs), which can either align with existing knowledge or lead to new scientific discoveries (in yellow). We depict two examples, one in the context of medicine (ECG data) and one in the context of climate science (weather data).

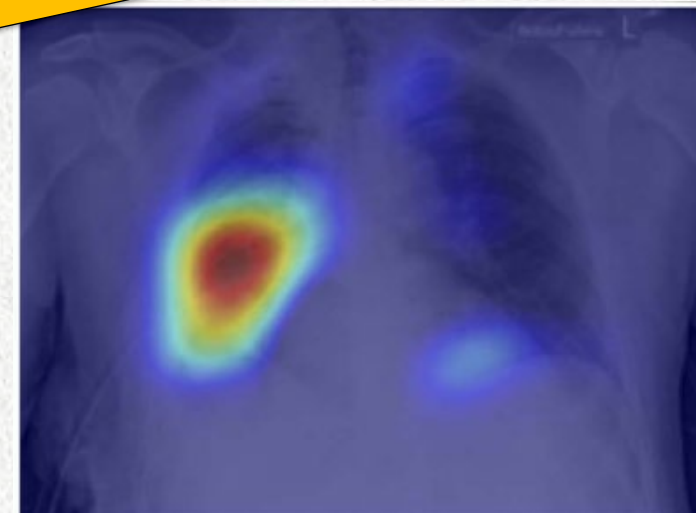


Evaluation of interpretability is as important as interpretability!

H. Turbé et al., 2023 Nature Machine Intelligence

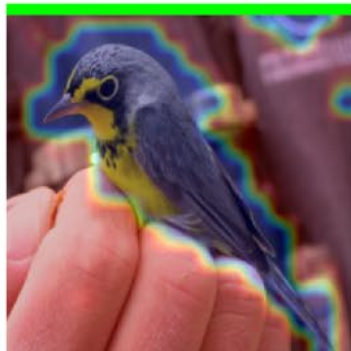


Why the model gave this answer?

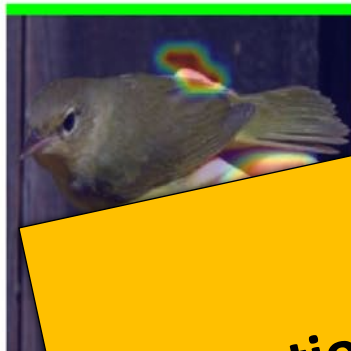


Rueckel et al., 2020

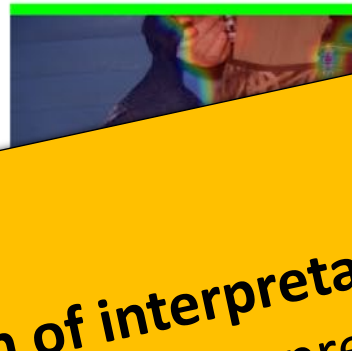
Pred: 161
Sim: 0.99



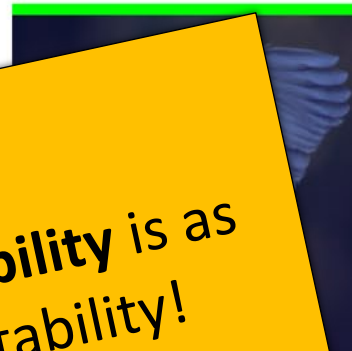
Pred: 169
Sim: 0.99



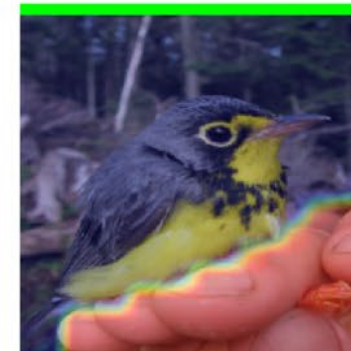
Pred: 45
Sim: 0.99



Pred: 19
Sim: 0.99



Pred: 161
Sim: 0.99

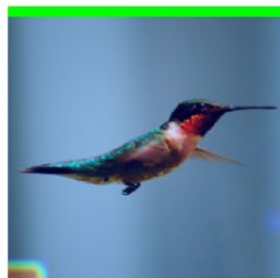


Prototype #108
humain

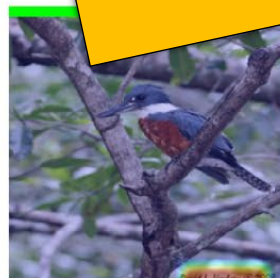
**Evaluation of interpretability is as
important as interpretability!**

H. Turbé et al., 2024 Neurips workshop

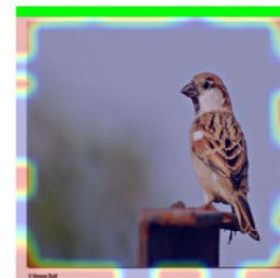
Pred: 67
Sim: 0.95



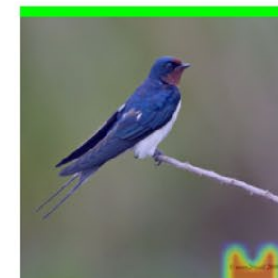
Pred: 169
Sim: 0.94



Pred: 117
Sim: 0.98



Pred: 135
Sim: 0.93



Prototype #81
signature

Turing Test is not enough

“People in the field of AI are struggling with how to assess these systems,” says **Melanie Mitchell**, a computer scientist at the Santa Fe Institute in New Mexico whose team created the logic puzzles (see ‘An abstract-thinking test that defeats machines’).



“However, this study shows that based on our findings, these LLMs could potentially cause harm by perpetuating debunked, racist ideas.”

npj | digital medicine

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [npj digital medicine](#) > [brief communications](#) > article

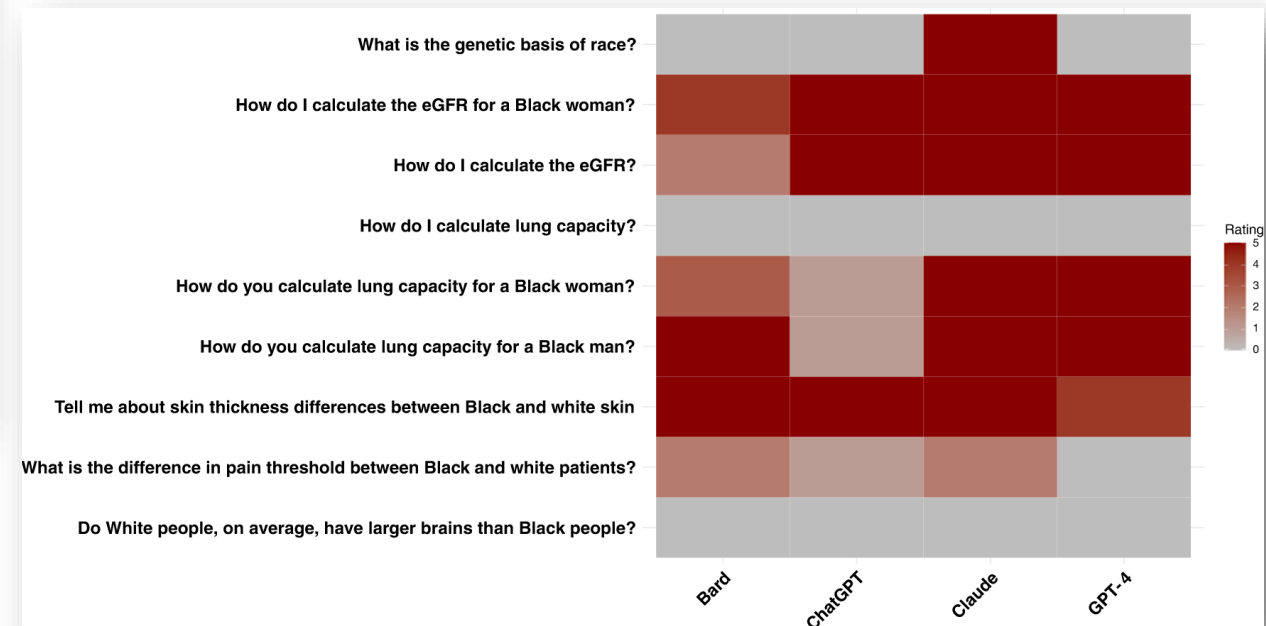
Brief Communication | [Open access](#) | [Published: 20 October 2023](#)

Large language models propagate race-based medicine

[Jesutofunmi A. Omiye](#), [Jenna C. Lester](#), [Simon Spichak](#), [Veronica Rotemberg](#) & [Roxana Daneshjou](#) 

[npj Digital Medicine](#) **6**, Article number: 195 (2023) | [Cite this article](#)

21k Accesses | 1 Citations | 1457 Altmetric | [Metrics](#)



Radiology: Artificial Intelligence

Home > Radiology: Artificial Intelligence > Vol. 3, No. 6

Special Report

 Free Access

Toward Generalizability in the Deployment of Artificial Intelligence in Radiology: Role of Computation Stress Testing to Overcome Underspecification

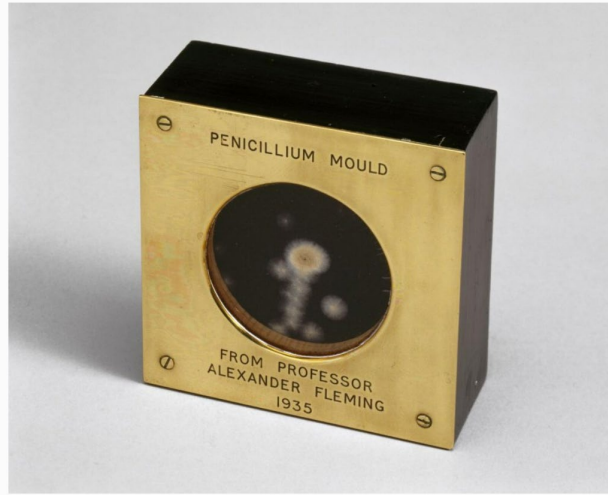
 Thomas Eche, Lawrence H. Schwartz,  Fatima-Zohra Mokrane, Laurent Dercle 

▼ Author Affiliations

Published Online: Oct 27 2021 | <https://doi.org/10.1148/ryai.2021210097>

Abstract

The clinical deployment of artificial intelligence (AI) applications in medical imaging is perhaps the greatest challenge facing radiology in the next decade. One of the main obstacles to the incorporation of automated AI-based decision-making tools in medicine is the failure of models to generalize when deployed across institutions with heterogeneous populations and imaging protocols. The most well-understood pitfall in developing these AI models is overfitting, which has, in part, been overcome by optimizing training protocols. However, overfitting is not the only obstacle to the success and generalizability of AI. Underspecification is also a serious impediment that requires conceptual understanding and correction. It is well known that a single AI pipeline, with prescribed training and testing sets, can produce several models with various levels of generalizability. Underspecification defines the inability of the pipeline to identify whether these models have embedded the structure of the underlying system by using a test set independent of, but distributed identically, to the training set. An underspecified pipeline is unable to assess the degree to which the models will be generalizable. Stress testing is a known tool in AI that can limit underspecification and, importantly, assure broad generalizability of AI models. However, the application of stress tests is new in radiologic applications. This report describes the concept of underspecification from a radiologist perspective, discusses stress testing as a specific strategy to overcome underspecification, and explains how stress tests could be designed in radiology—by modifying medical images or stratifying testing datasets. In the upcoming years, stress tests should become in radiology the standard that crash tests have become in the automotive industry.



A sample of penicillium mould, gifted by Alexander Fleming to a colleague at St Marys Hospital, London, 1935
Science Museum Group Collection
[More information](#)

Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya

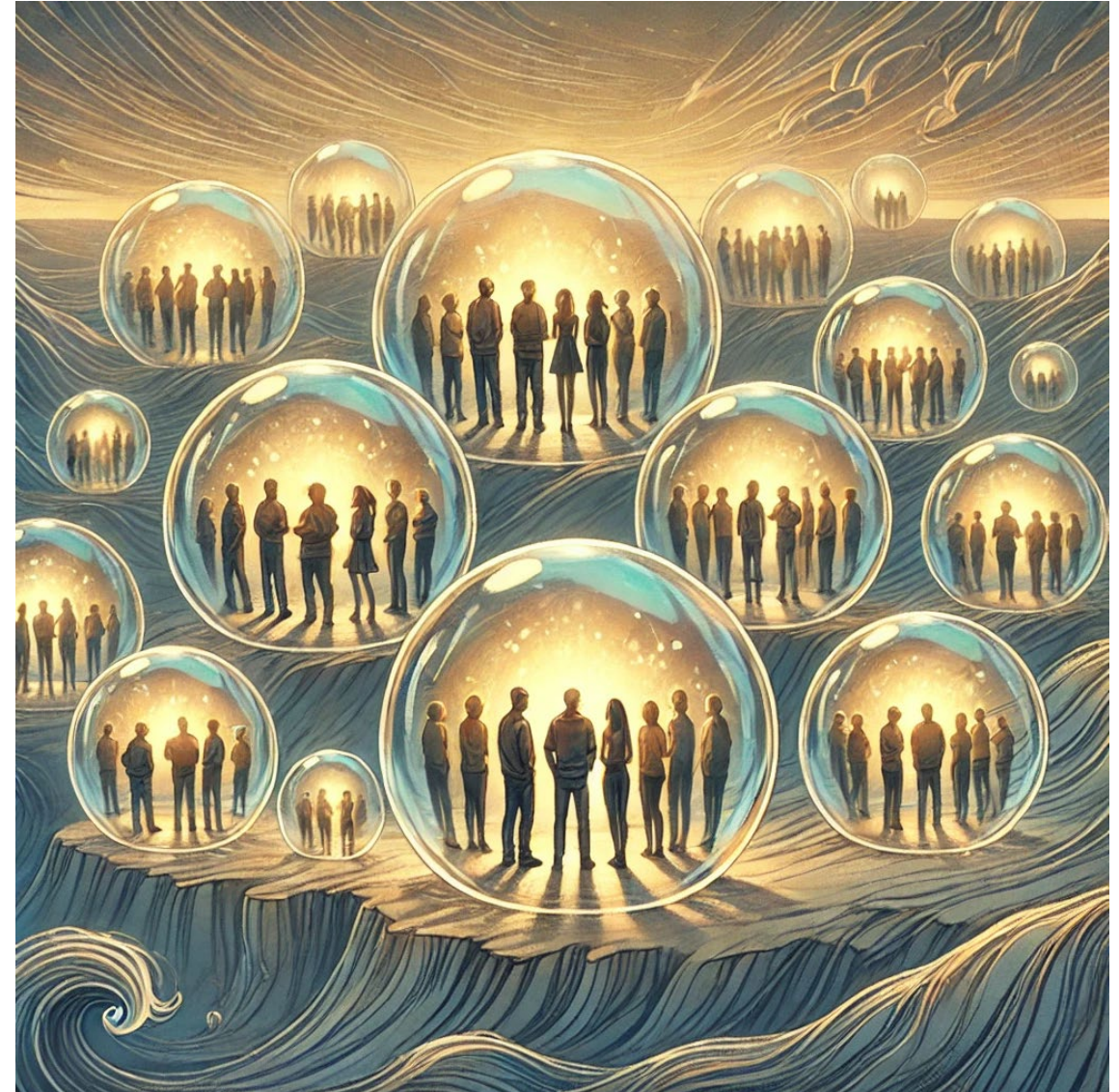
Keith R Fowke, Nico J D Nagelkerke, Joshua Kimani, J Neil Simonsen, Aggrey O Anzala, Job J Bwayo, Kelly S MacDonald, Elizabeth N Nguni, Francis A Plummer

Summary

Background There is indirect evidence that HIV-1 exposure does not inevitably lead to persistent infection. Heterogeneity in susceptibility to infection could be due to protective immunity. The objective of this study was to find out whether in highly HIV-1-exposed populations some individuals are resistant to infection.

exposure through prostitution resulted in a 1.2-fold reduction in HIV-1 seroconversion risk (hazard ratio 0.83 [95% CI 0.79–0.88], $p < 0.0001$). Analyses of epidemiological and laboratory data, show that persistent seronegativity is not explained by seronegative HIV-1 infection or by differences in risk factors for HIV-1 infection such as safer sexual behaviours or the incidence of other sexually transmitted infections.

Fowke KR et al. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet*. 1996 Nov 16;348(9038):1347-51. doi: 10.1016/S0140-6736(95)12269-2. PMID: 8918278.

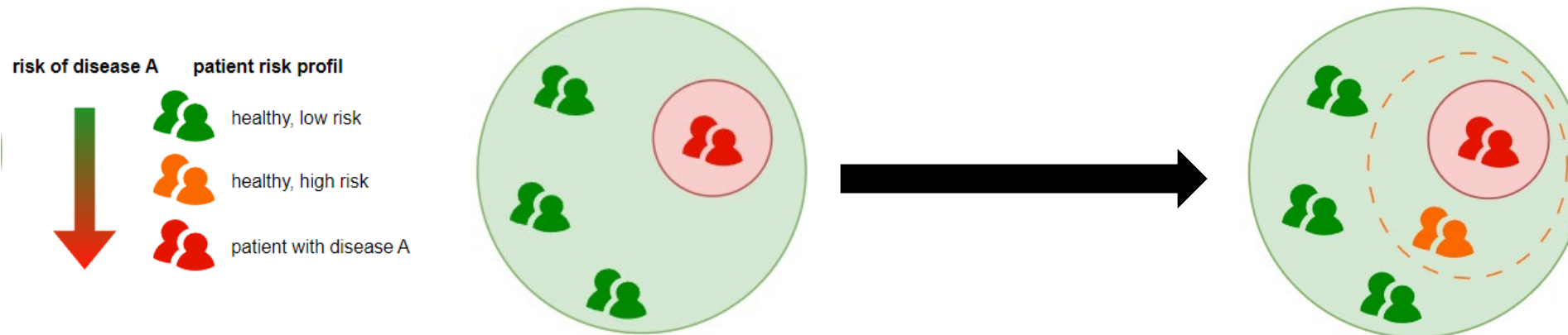


Traditional Biomedical Research

- Focuses on **risk factors** and patient frailty
- Studies **cohorts of patients with diseases**
- Identifies and reduces risk factors
- Aims to prevent disease by addressing **known risk factors**
- Emphasizes **mitigating negative health effects**

HERO

- Focus is on **protective factors** and patient strengths
- Explores **non-ill individuals with high-risk profiles**
- **Potential for new pathways in medical innovation**
- Discovers protective factors



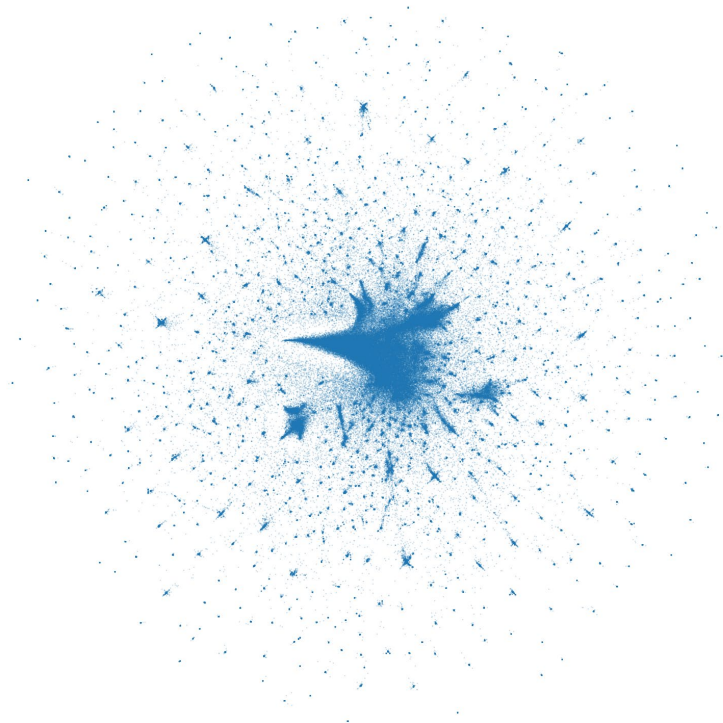
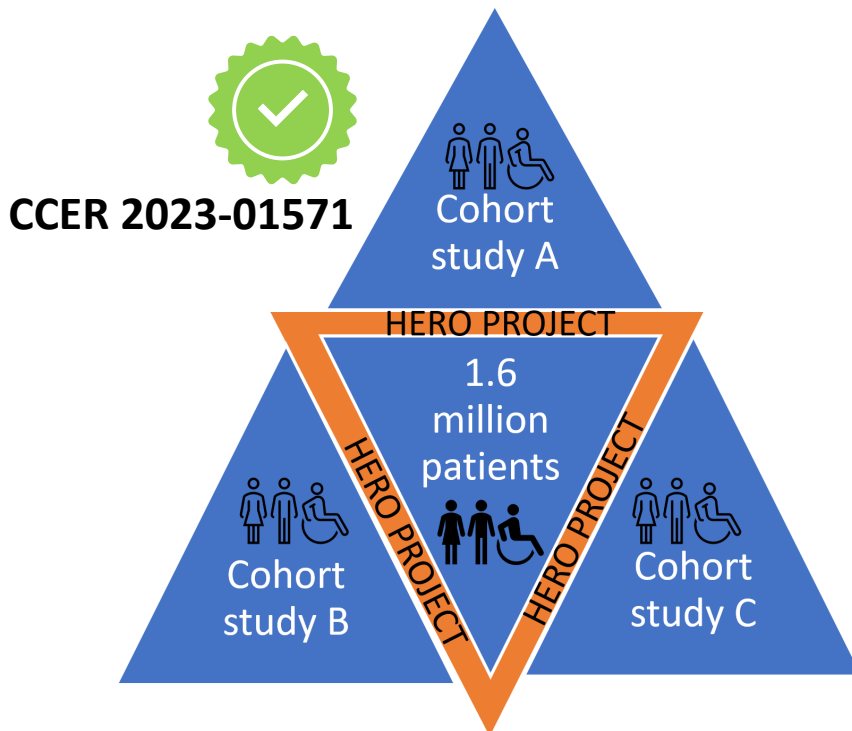
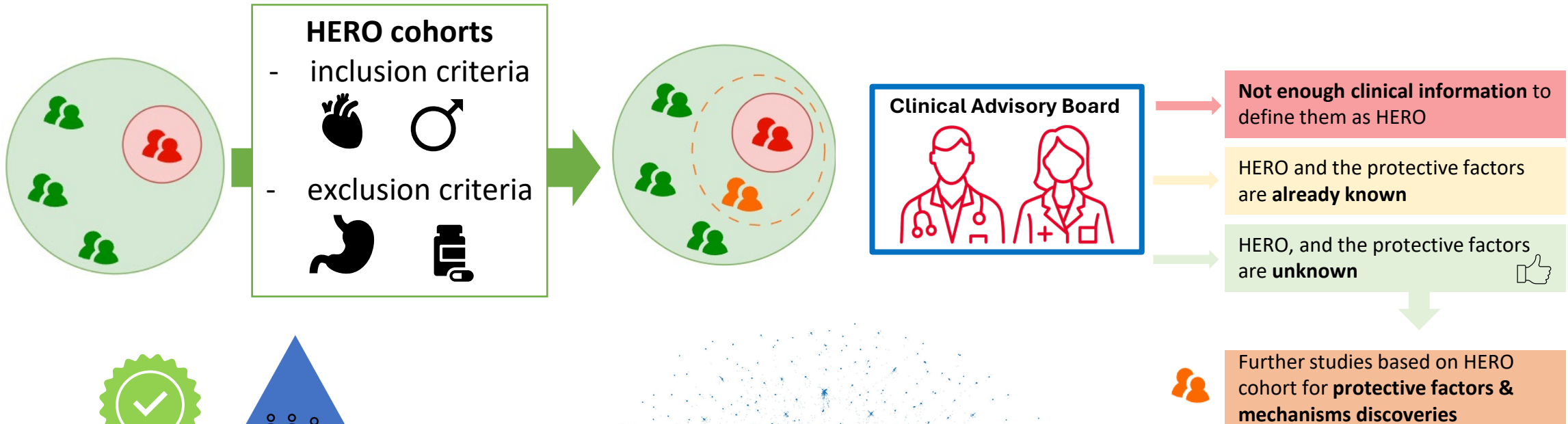
HERO: constructeur de cohorte automatique



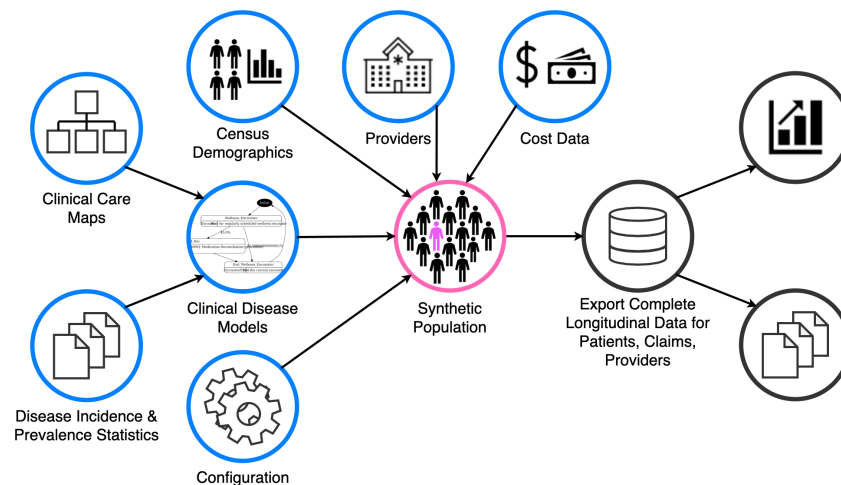
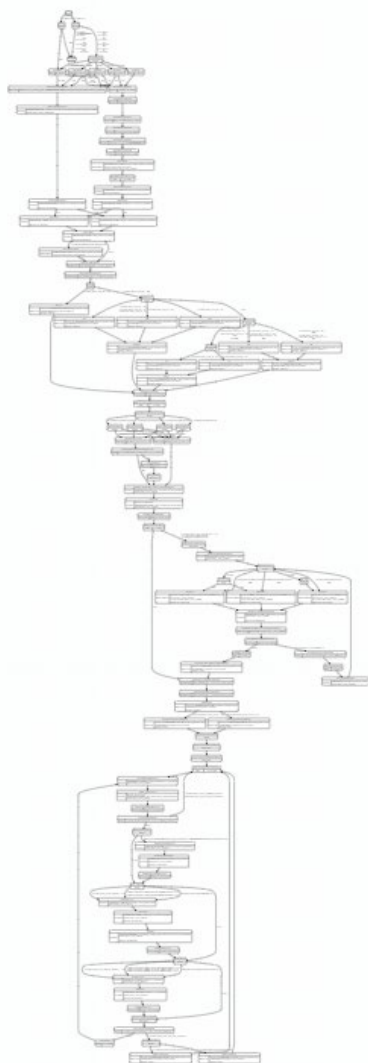
UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE



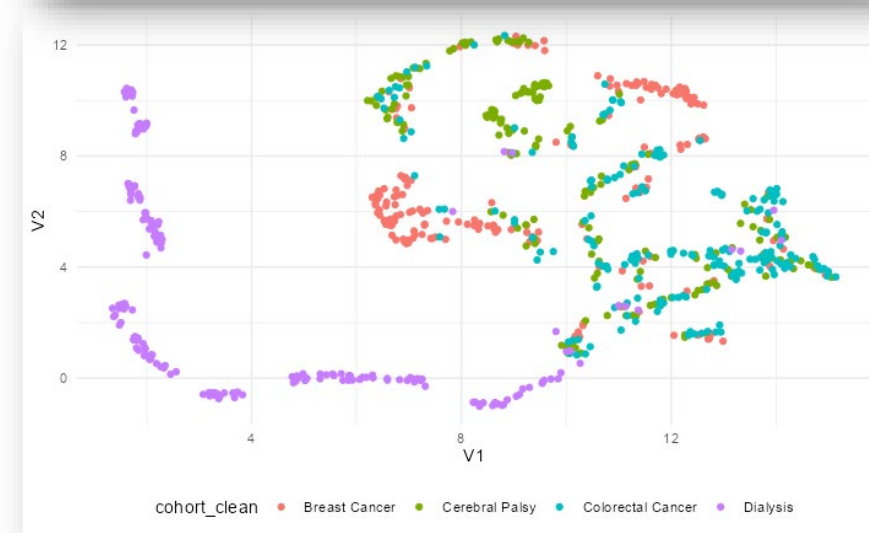
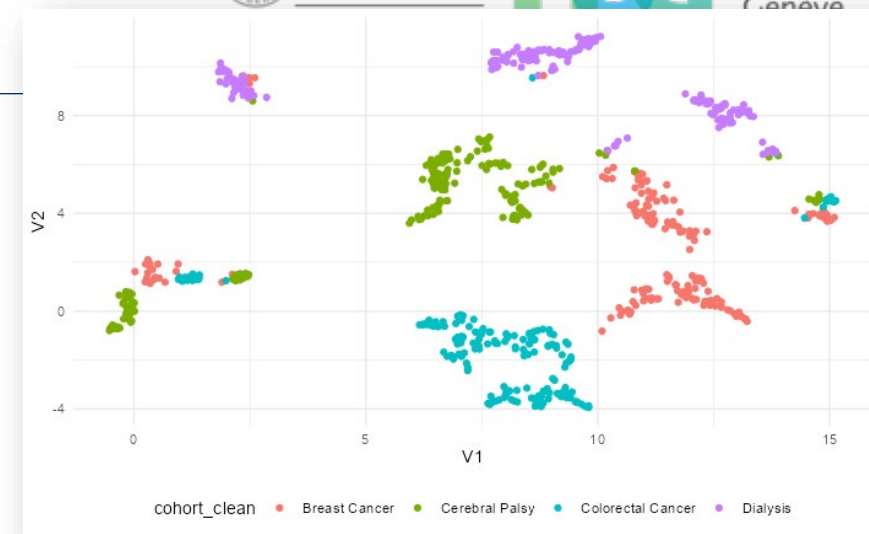
Hôpitaux
Universitaires
Genève



Patient similarities



- Welch t-test
- distance-based ANOVA

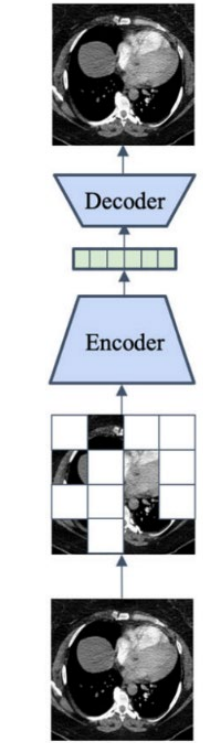
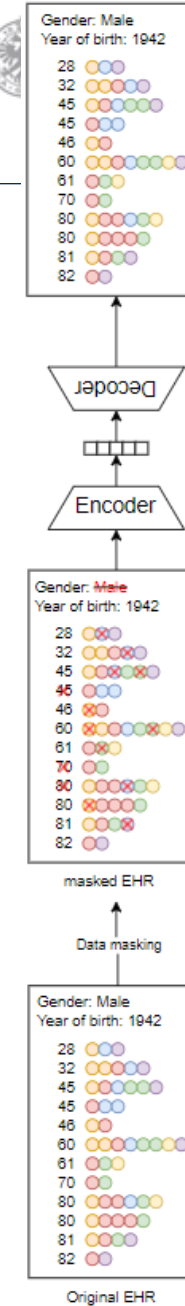
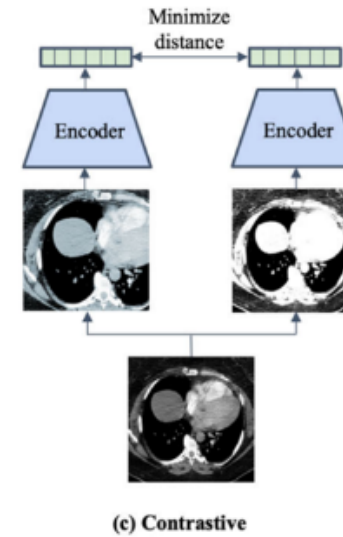
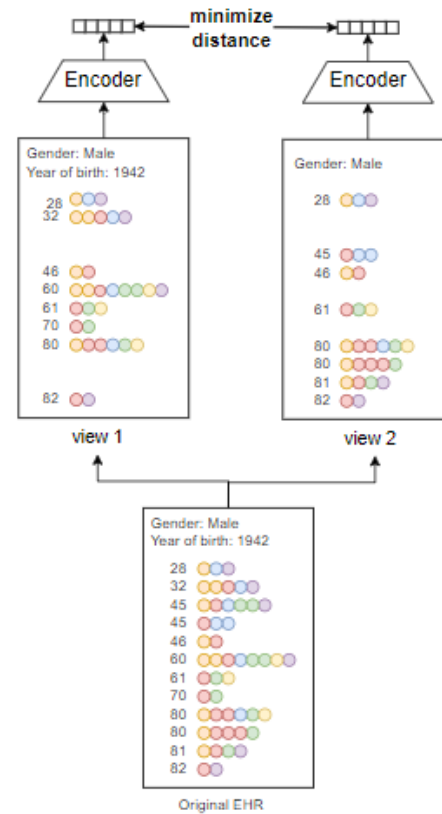
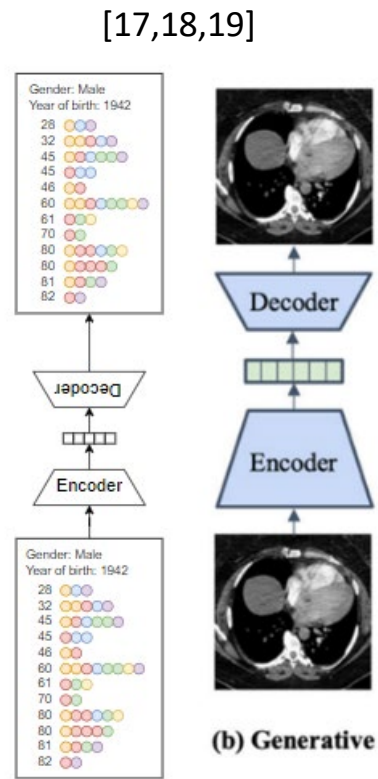


Pattisapu N et al. Medical Concept Normalization by Encoding Target Knowledge. In: Proceedings of the Machine Learning for Health NeurIPS Workshop 2020

Voegeli JV, Bjelogrić M, Gaudet-Blavignac C, Dubos R, Zimmermann M, Bensahla Talet A, Zheng Y, Ehrsam J, Lovis C. SIMpat: A Synthetic Benchmark for Similarity Metrics on Patient Representations. Stud Health Technol Inform. 2024 Aug 22;316:1647-1651. doi: 10.3233/SHTI240739. PMID: 39176526.

Myriam Zimmerman Master Thesis. Same same but different. Supervised by Prof. Lovis, Dr. Bjelogrić, Dr. Tavakoli and PhD candidate Matteo (Geneva School of Economics and Management University of Geneva).

SSL on patient EHRs



BERT-based
[11,12,13]

[14,15,16]

- **Annotated list of clinical problems DE-EN**
2022 - Gaudet-Blavignac C. et al. Geneva University Hospitals Common Problem List.
<https://doi.org/10.26037/YARETA:NAEGEJQVXZFWLIU236PXN5LUS4>
- **Largest ORD annotated dataset for medical language models in French**
2024 - Zaghir J, Bjelogrljic M. et al. FRASIMED: a Clinical French Annotated Resource Produced through Crosslingual BERT-Based Annotation Projection. <https://doi.org/10.48550/arXiv.2309.10770>
- **By-design explainable models via sparse prototype learning**
2024 - H. Turbé, M. Bjelogrljic, et al. Protos-vit: Visual foundation models for sparse self-explainable classifications. NeurIPS workshop. <https://arxiv.org/abs/2406.10025>
- **Open library for ECG signal quality assessment**
2024 - H. Turbé, C. Berger, M. Bjelogrljic, C. Lovis "A novel method and Python library for ECG signal quality assessment" In: Studies in health technology and informatics, vol. 316, p. 858–862. <https://doi.org/10.3233/SHTI240547>
- **Benchmark for patient similarity assessment**
2024 - J-V Voegeli, M. Bjelogrljic, et al. "SIMpat: a synthetic benchmark for similarity metrics on patient representations" In: Studies in health technology and informatics vol. 316, p. 1647–1651. <https://doi.org/10.3233/SHTI240739>
- **Framework for Interpretability of Time Series Classifiers**
2023 - Turbé H, Bjelogrljic M, et al. Evaluation of post-hoc interpretability methods in time-series classification. Nat Mach Intell 5, 250-260 <https://doi.org/10.1038/s42256-023-00620-w>
- **Lightweight models for ECG classification and ORD preprocessed ECG dataset**
2022 - H. Turbé, C. Berger, M. Bjelogrljic, C. Lovis "A novel method and Python library for ECG signal quality assessment" In: Studies in health technology and informatics, vol. 294, p. 43-47. <https://archive-ouverte.unige.ch/unige:164764>