RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inserm
La science pour la santé
From science to health

IA et numérique pour la recherche en santé

Hugues Berry

Pôle IA & Numérique, Inserm

hugues.berry@inserm.fr pole.ia@inserm.fr

Plan

- **IA: de quoi parle-t-on?**
- **Les obstacles principaux**
- **Le Pôle IA et Numérique de l'Inserm**

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inserm
La science pour la santé
From science to health

IA: de quoi parle-t-on?

Ca fait quoi, l'IA (Machine & Deep Learning)?

Trois méthodes de base

Classification / Regression

Embedding

Estimation des distributions

Classification et regression

La tâche à effectuer

absence patho

présence patho

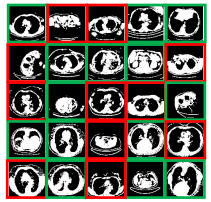
score patho ∈ [0.0,...,1.0]

Apprentissage supervisé (annotations)

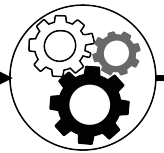
`random forests`
`gradient boosting`

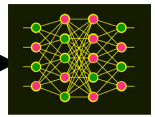
`MLPs`
`RNNs`

Ground Truth

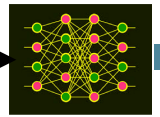


Training





Inference



absence patho

présence patho

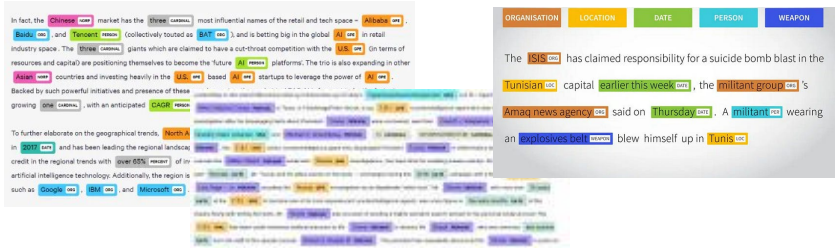
Classification et regression

La segmentation comme une tâche de classification des pixels

CT slice | Ground truth | Overlapping

`U-Nets, CNNs`

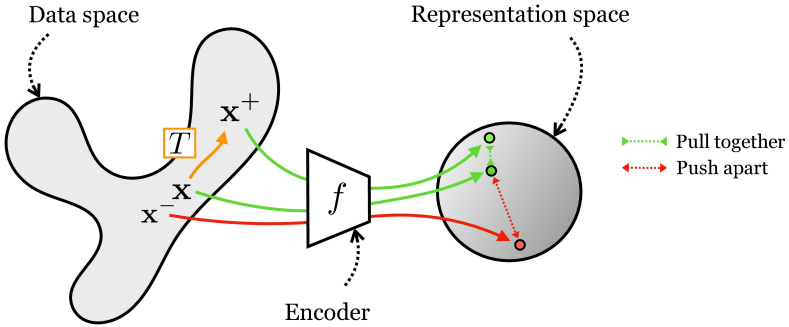
Le TAL comme une tâche de classification des mots/phrases

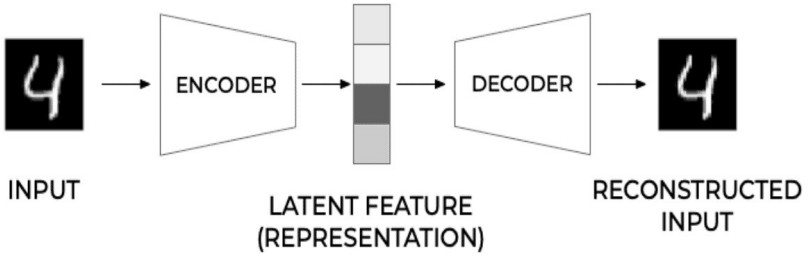


`Transformers (BERT)`
`CamemBERT`
`BioBERT`
`PubmedBERT`
`...`

Embedding / espace latent

Apprentissage des représentations

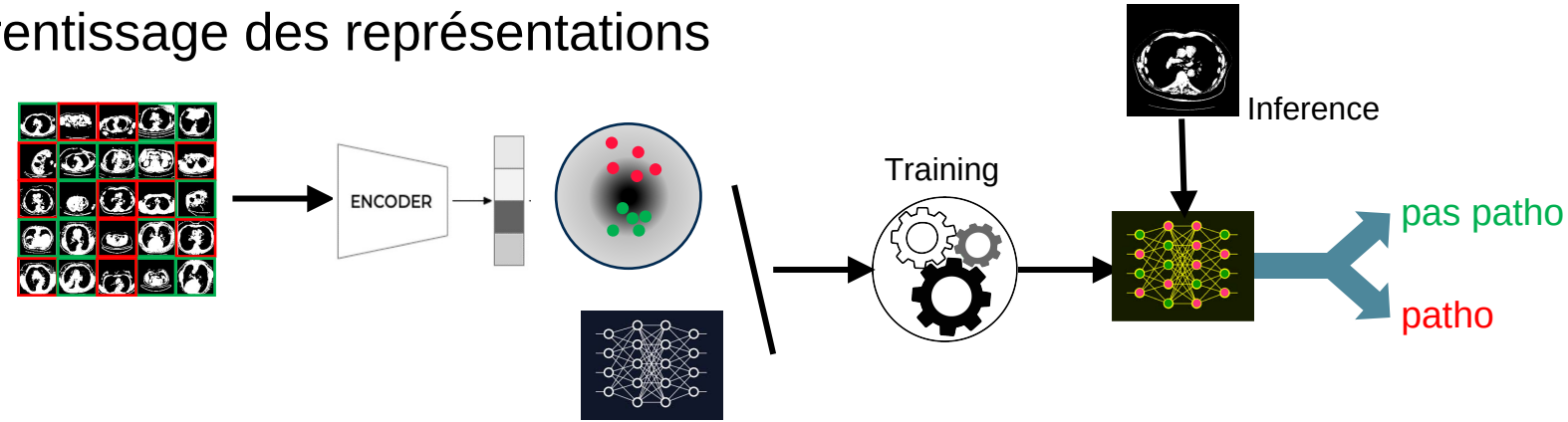




`AutoEncoder`

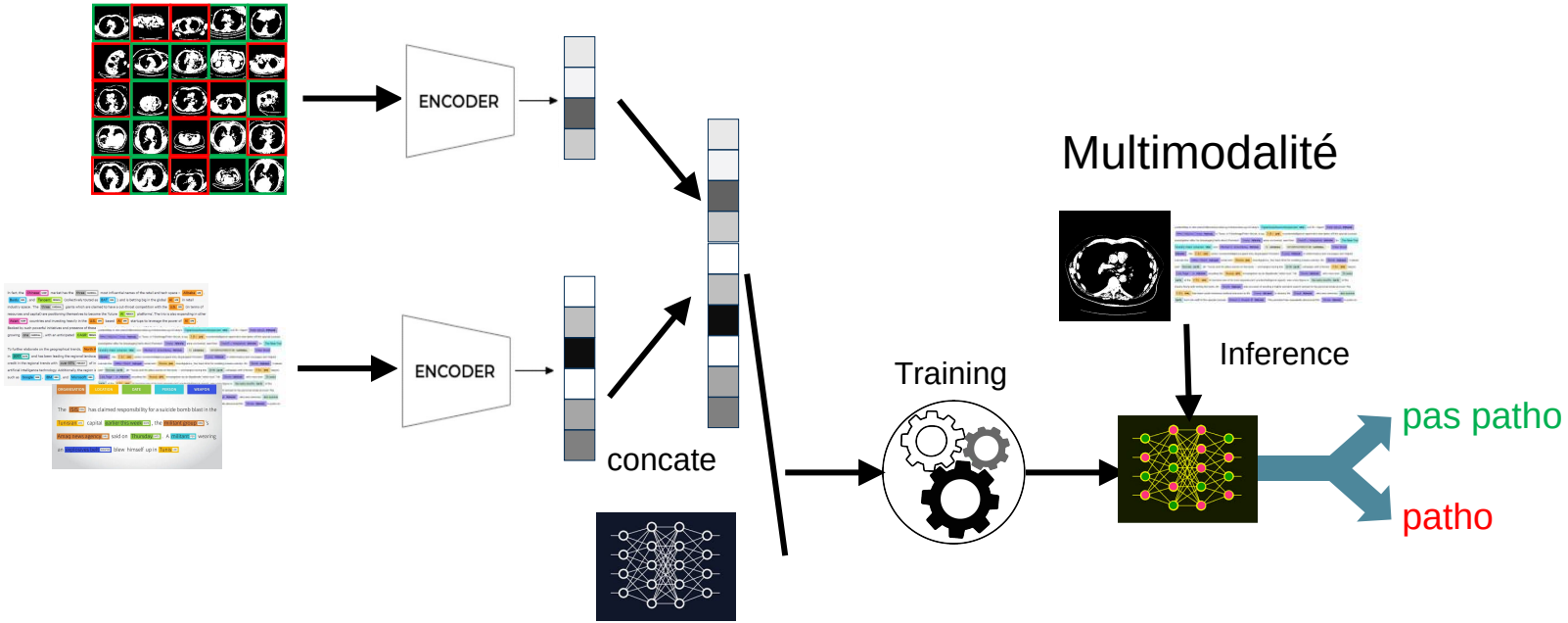
Embedding / espace latent

Apprentissage des représentations

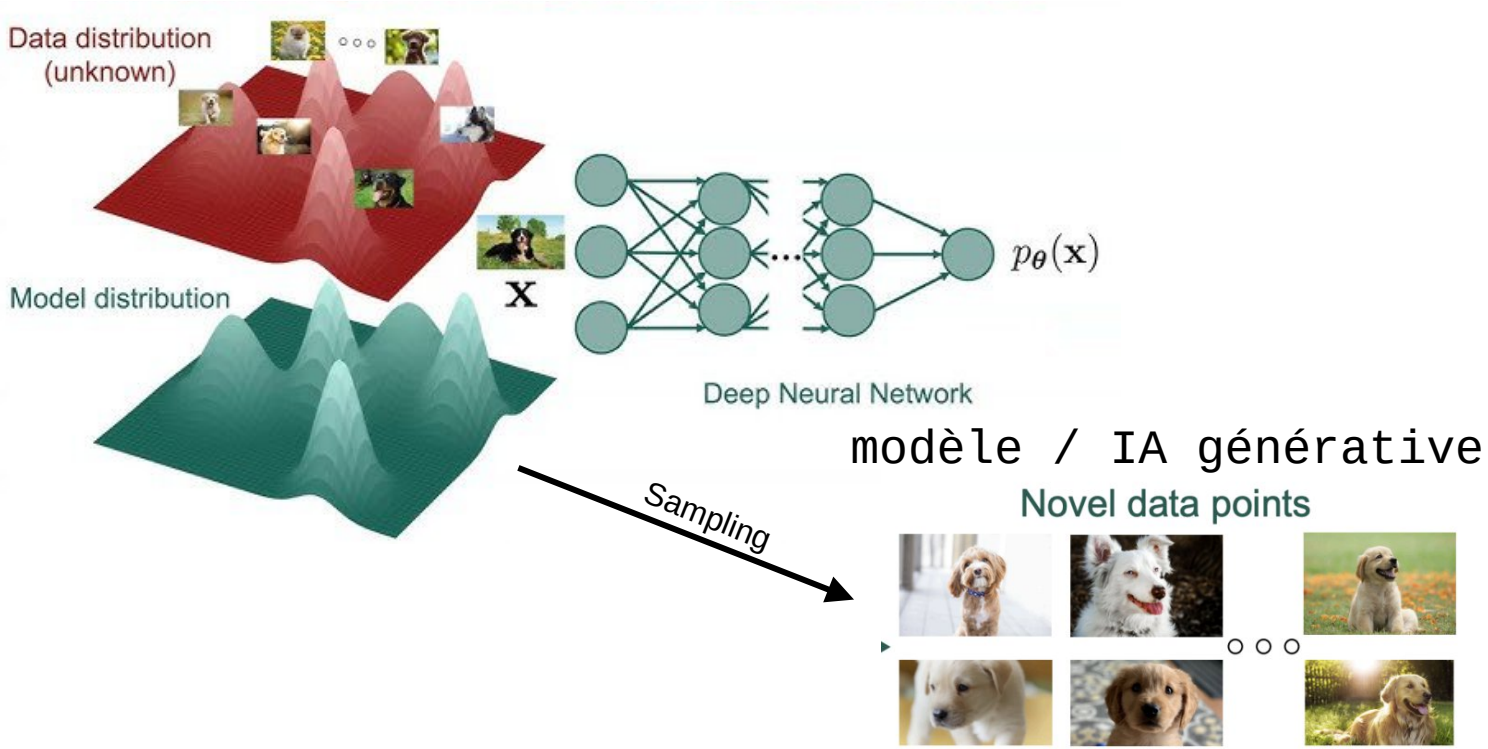


Embedding / espace latent

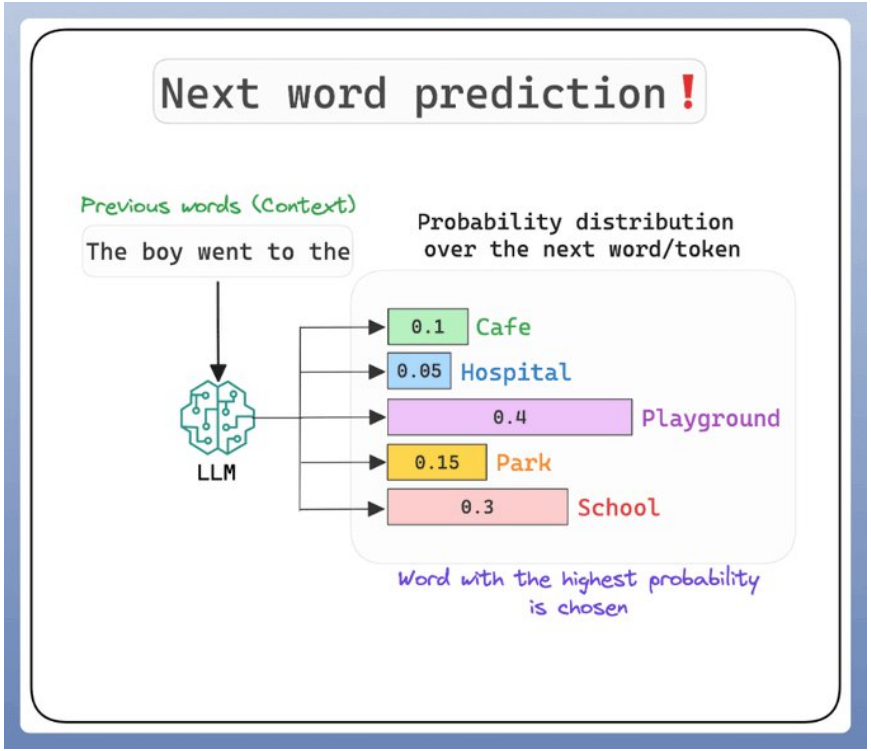
Apprentissage des représentations



Inférence des distributions



Inférence des distributions: les LLMs



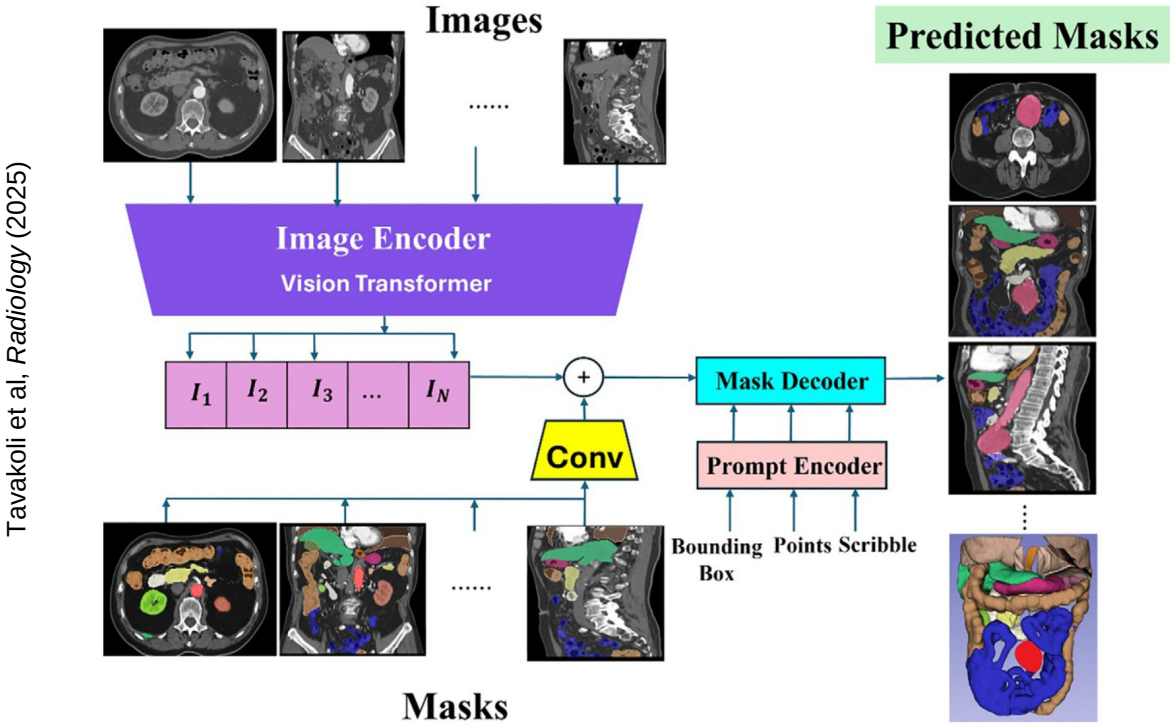
modèles de langage génératifs (LLM)

GPT
Claude
Gemini
MistralAI
...

+ specific training for Q&A tasks

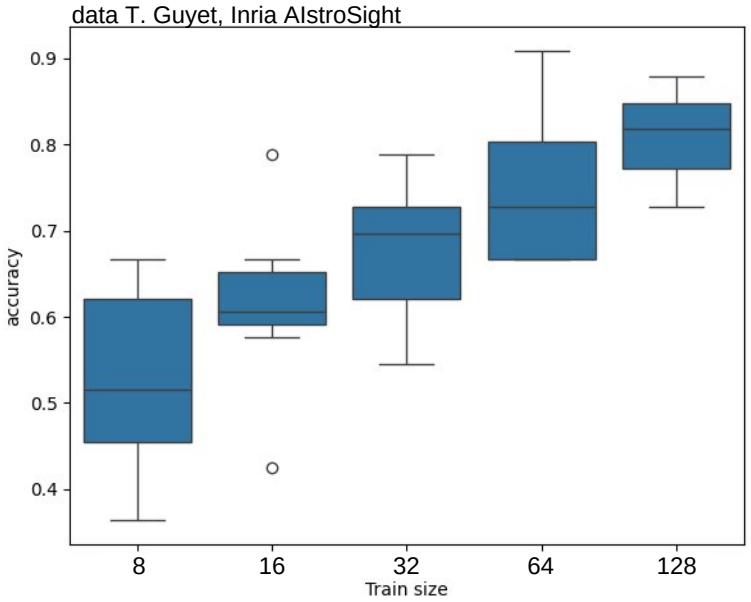
chatGPT
chatbots

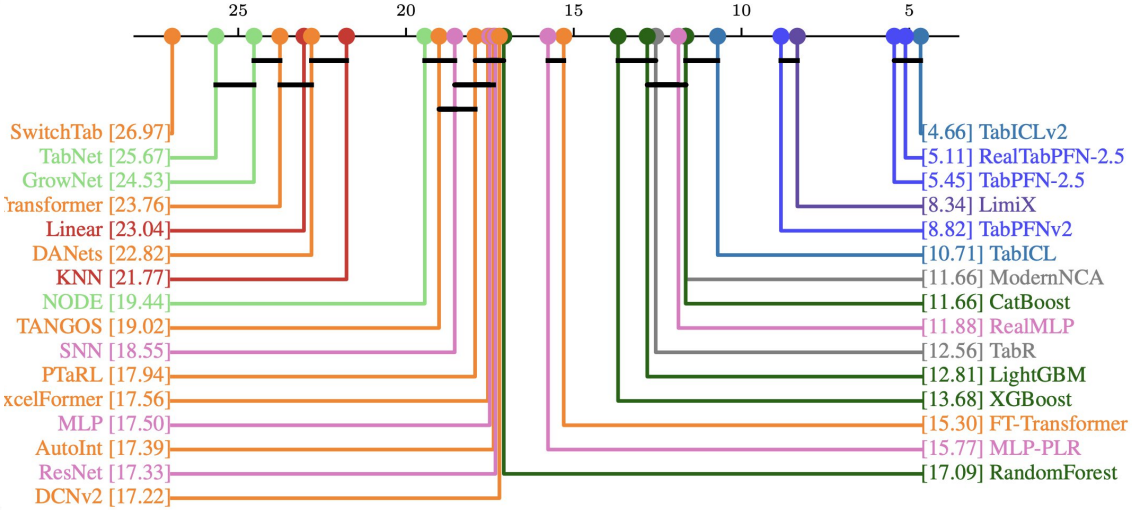
Généralisation des LLMs: les modèles de fondation



Tavakoli et al, *Radiology* (2025)

Les modèles de fondation pour données tabulaires



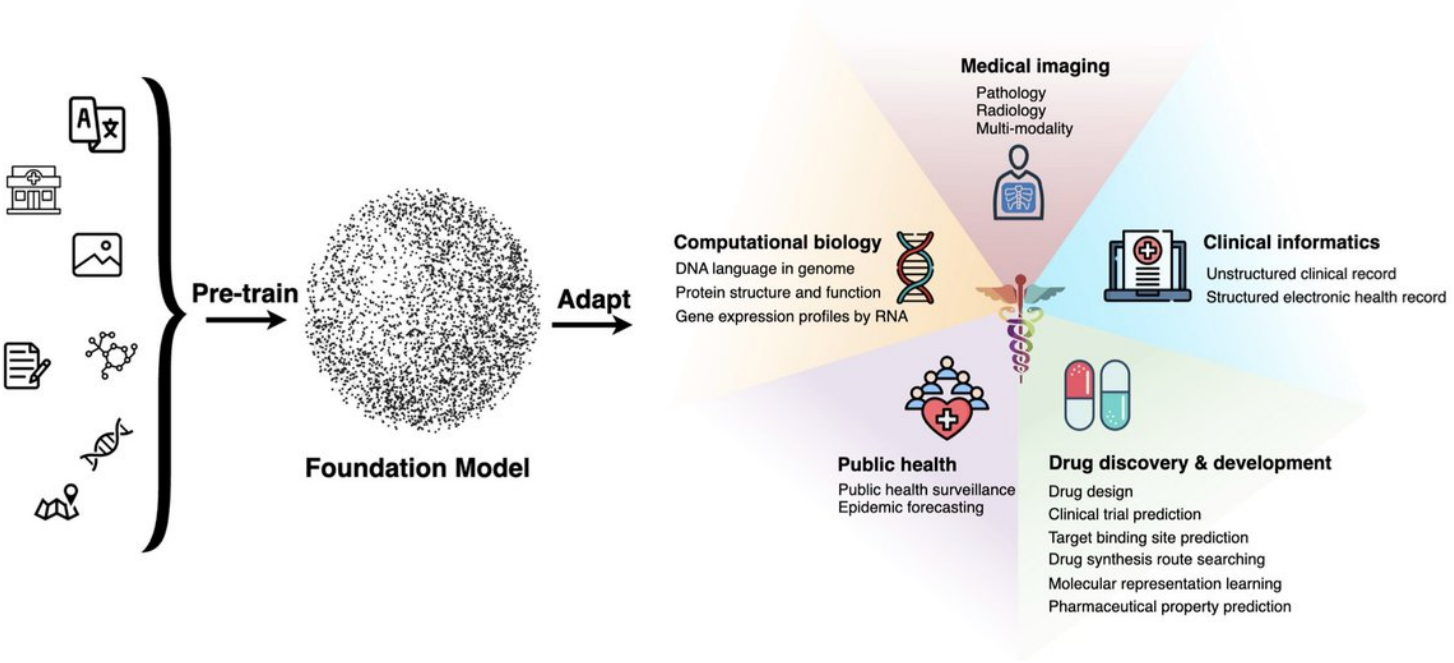


Classement moyen (accuracy) sur 300 datasets (classif +regression)

Qu et al, 2026. hal-05538427

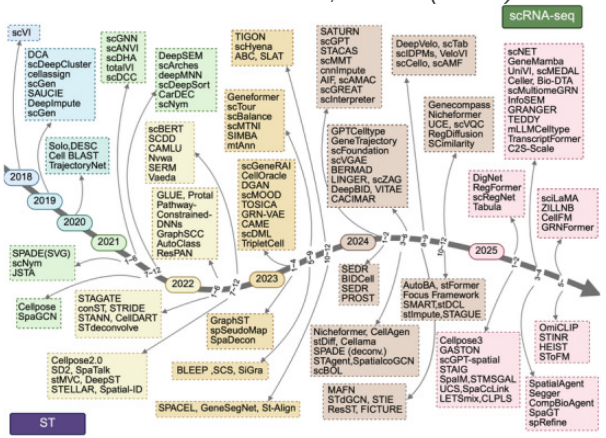
Des modèles de fondation pour toute la biomédecine

Liu et al, *arXiv*:2503.02104 (2025)



AI for Bioinformatics: a focus on single-cell omics

Li et al, *Adv Sci (Weinh)* 2025



Batch effect correction

BERMAD (Zhan et al, *Bioinform* 2024); scBCN (Wan et al, *Briefing Bioinform* 2025)

Dimensionality reduction and clustering

scDeepCluster (Tian et al, *Nat Mach Intell* 2019), scGPT (Cui et al, *Nat Meth* 2024)

Cell type annotation

scBERT (Yang et al, *Nat Mach Intell* 2022), CANAL (Wan et al, *Brief Bioinform* 2024)

Differentiation trajectories (Optimal Transport)

TIGON (She et al, *Nat Mach Intell* 2023), STORIES (Huizing et al, *Nat Meth* 2025)

Perturbation prediction

GEARS (Roohani et al, *Nat Biotech* 2023), scOTM (Wang et al, *Bioengineering* 2025)

Multimodal integration

scConfluence (Samaran et al, *Nat Commun* 2024), scMODAL (Wang et al, *Nat Commun* 2025)

Inference of gene regulatory networks

scPRINT (Kalfon et al, *Nat Commun* 2025), LINGER (Yuan & Duren *Nat Biotechnol* 2025)

Many foundation models that do (almost) all this

scFoundation (Hao et al, *Nat Meth* 2024), scGPT (Cui et al, *Nat Meth* 2024), CellFM (Zeng et al *Nat Commun* 2025)...

for sc and spatial transcriptomics: Tchatchoua Ngassam et al, *Front Bioinform* 2026

L'IA comme assistant co-scientist

Formulation d'hypothèses par LLM / aide à la découverte

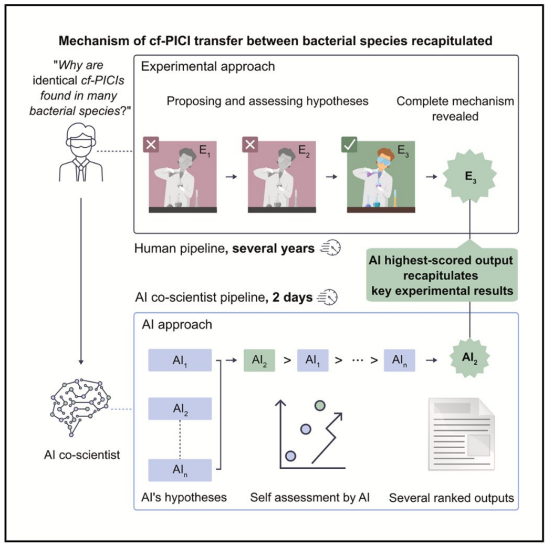
Pénadès et al. *Cell* (2025)

Cell

Theory

AI mirrors experimental science to uncover a mechanism of gene transfer crucial to bacterial evolution

Graphical abstract



Authors

José R. Penadés, Juraj Gottweis, Lingchen He, ..., Vivek Natarajan, Alan Karthikesalingam, Tiago R.D. Costa

Correspondence

j.penades@imperial.ac.uk (J.R.P.), natviv@google.com (V.N.), alankarthi@google.com (A.K.), t.costa@imperial.ac.uk (T.R.D.C.)

In brief

By solving a previously unsolved biological question, the AI co-scientist predicted a complex mechanism of gene transfer and generated hypotheses that opened new research directions, illustrating AI's potential as a creative engine in discovery.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inserm
La science pour la santé
From science to health

Les obstacles principaux

L'IA a besoin de beaucoup de données

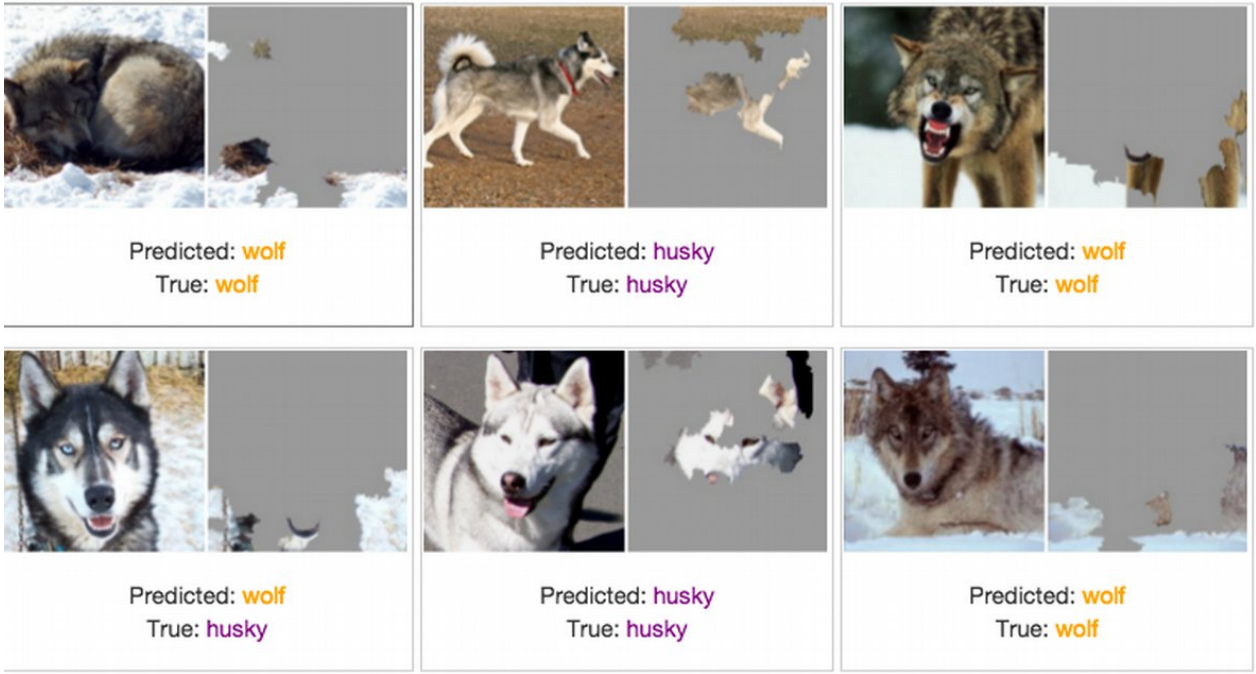
Une tension entre deux necessitiés:

partager les data (augmenter volume)

tout en assurant anonymat patients

et en maintenant contraintes en termes de publication

Le probleme du biais dans les données



Ribeiro et al, *KDD* 2016

Le probleme du biais dans les données

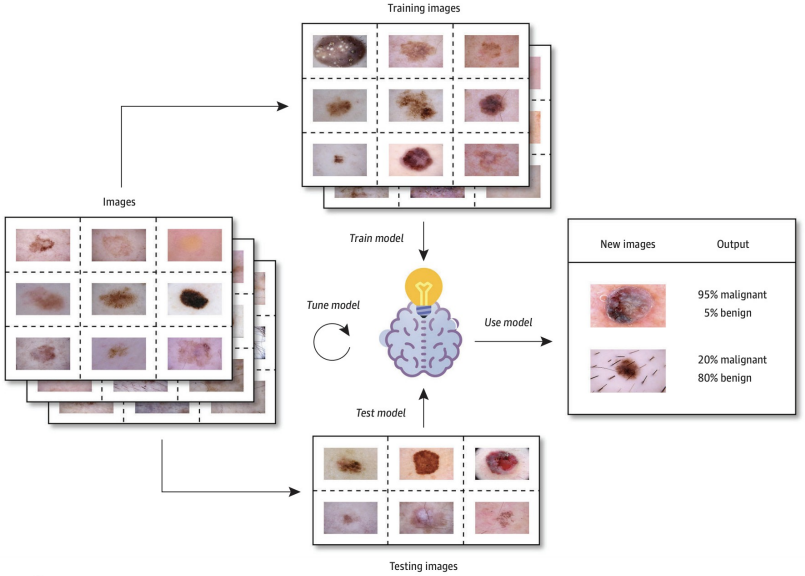
Opinion

VIEWPOINT

Machine Learning and Health Care Disparities in Dermatology

Adewole S. Adamson, MD, MPP
Department of Dermatology, University of North

Machine learning (ML), a form of artificial int
using computer algorithms, is often applied ir
take for granted: Spotify to predict music th
may enjoy, Facebook to suggest friends to tag i



nize disease. Unfortunately, most ML programs are largely learning on light skin. For example, in the International Skin Imaging Collaboration: Melanoma Project, which is one of the largest and often-used, open-source, public-access archives of pigmented lesions, much of the patient data are heavily collected from fair-skinned populations in the United States, Europe, and Australia.[3] Thus, no matter how advanced the ML algorithm, it may underperform on images of lesions in skin of color.

embraced by dermatologists. However, this new technology also has the potential to leave marginalized groups behind. Although most early ML studies in dermatology focused on pigmented lesions, ML may eventually be developed for evaluating rashes, which can visually manifest

Adamson & Smith, *JAMA Dermatol*, 2018

Les systèmes d'IA gourmands++ en ressources

Apprentissage = énormes data centers

data center moderne = 1.0-1.4 GW

= production d'1 réacteur d'une centrale nucléaire française (0.9-1.6 GW)

St. Joseph County, Indiana, 2.2 GW

Même tension sur conso d'eau

data center = jusqu'à 19 million litres d'eau chaque jour

= conso d'une ville de 50,000 ie Passau or Vincennes

Dublin, Ireland

Keywords: frugalité, (in)formation

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

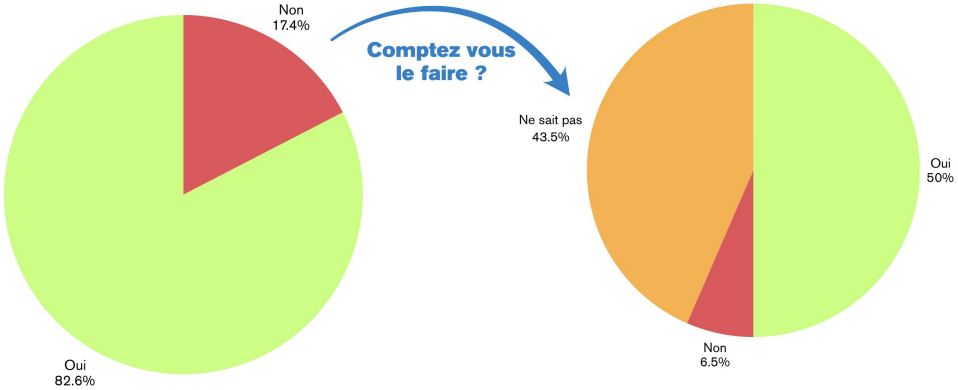
Inserm
La science pour la santé
From science to health

Le Pôle IA et Numérique de l'Inserm

Un début de cartographie (sondage sept 2025)

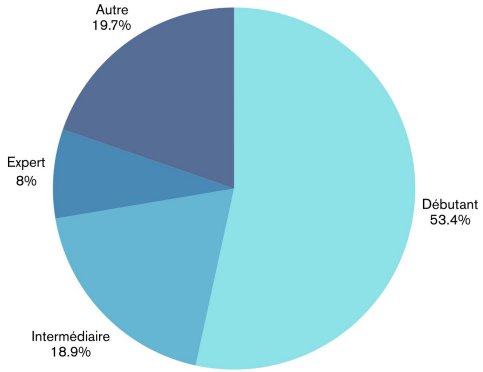
265 réponses (133 unités = 38%)

Utilisez-vous l'IA ?



90% utilisent ou prévoient d'utiliser l'IA

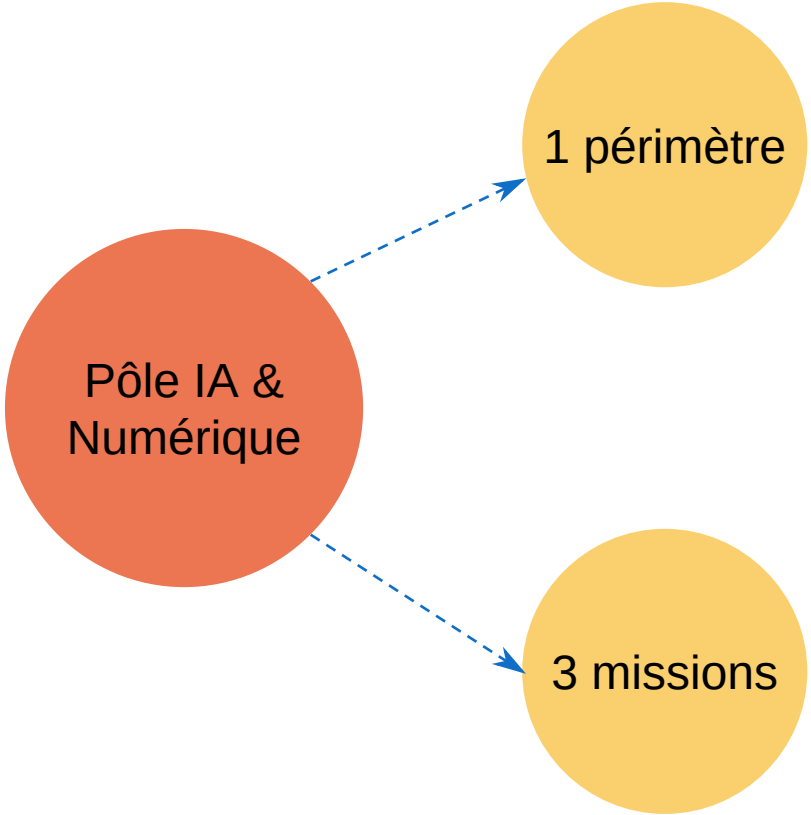
Quel est votre niveau de maturité ?



Détail de autre :
Souvent des équipes mixtes où il peut à la fois y avoir des experts et des débutants

>70% manquent d'expertise

analyse complète (~15 questions) : pole.ia@inserm.fr



L'IA & le Numérique pour la recherche

Lien transverse pour stratégie Inserm

Emission de recommandations

Accompagner montée en capacité

Aider à la montée en capacité

Augmenter la présence de numéricien.ne.s

Centres de ressources IA et Données

- guichet conseil
- séjours pour developpement dans unités
- animation locale IA - santé
- mutualisable
- lieux de formation interne

è 1 premier en phase d'expérimentation

è début de positionnement sur d'autres sites

è objectif: 10 Centres (COMP)

généré par Gemini

Aider à la montée en capacité

Augmenter la présence de numéricien.ne.s

+ de stagiaires écoles ingénieur IA-Santé

accords UTC, IMT-A, Insa Lyon, EPITA, Telecom Paris...

è premiers match-making en 2025

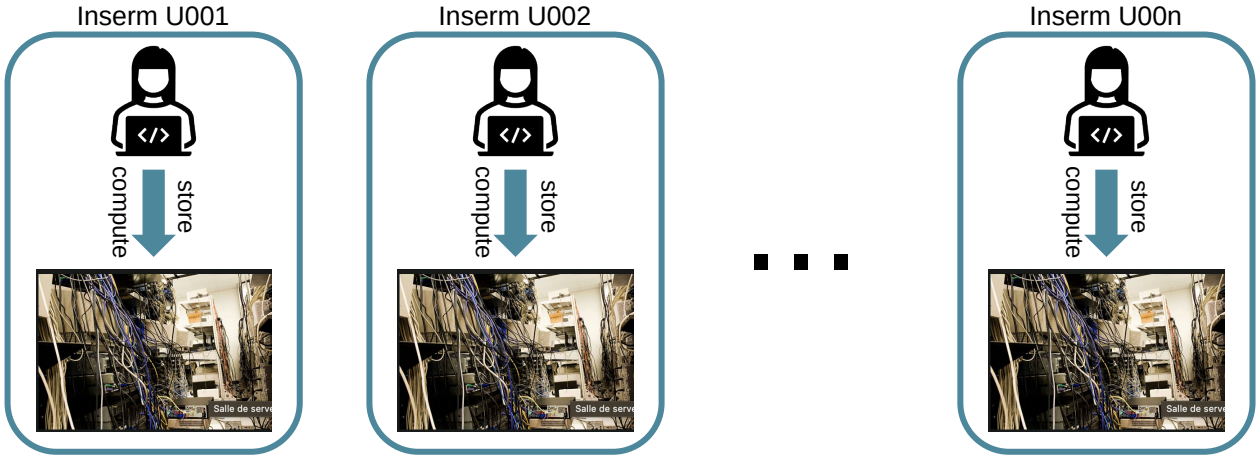
è elargissement international en 2026:

- è MoU avec U Concordia (Montreal)
- è projections UCL, U Laval, Inde

généré par Gemini

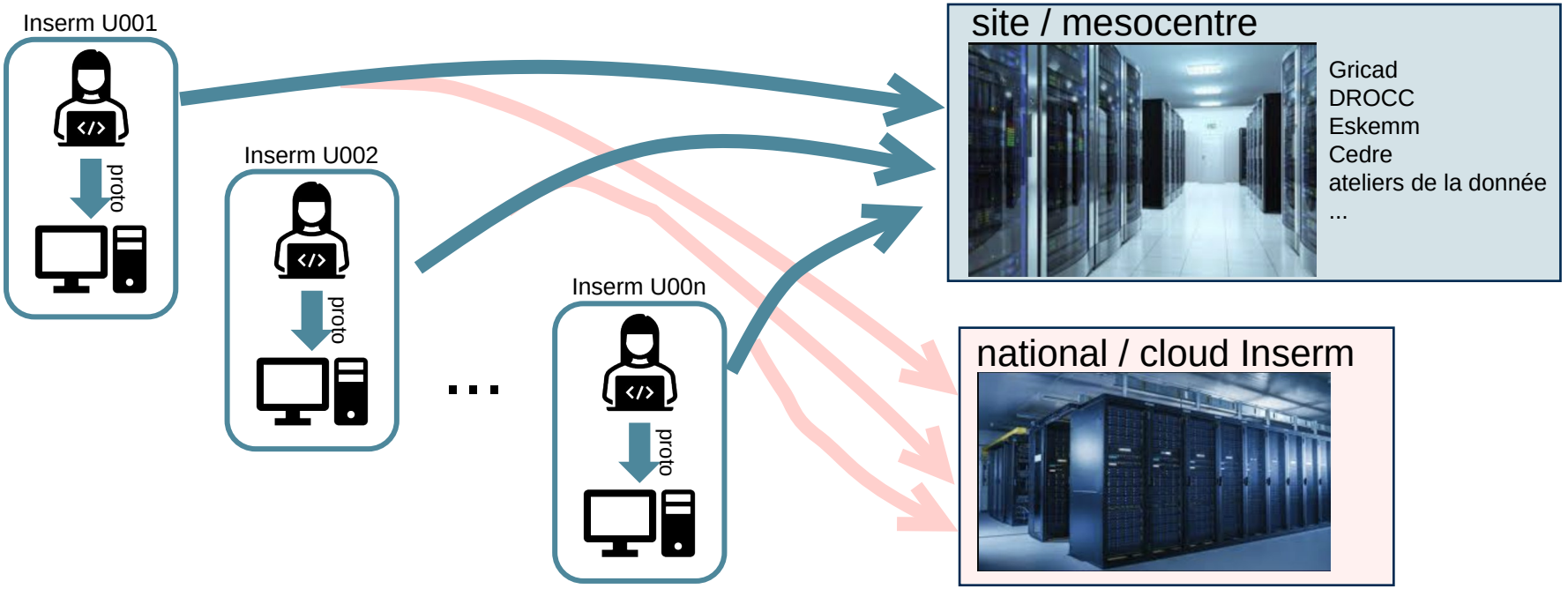
Stratégie en termes d'infrastructure calcul et stockage

- Fin du microlocal, priorité au site, débordement au national



Stratégie en termes d'infrastructure calcul et stockage

- Fin du microlocal, priorité au site, débordement au national



Aider à la montée en capacité: moyens numériques

Le cloud Inserm:

on premise, souverain

HDS et non-HDS

infra AWS-like (vmware)

Services IA dans le cloud Inserm

- IAg pour Q&A, RAG, MCP
- multi-agentic en mode project ou en autonomie
- HPC / slurm pour IA scientifique

En attendant les outils Inserm on-prem:

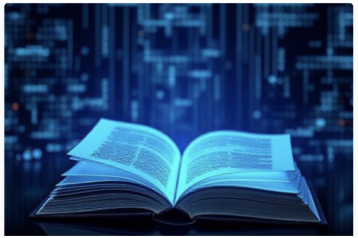
- Stockage des données non de santé: au delà de share.inserm.fr *recherche.data.gouv.fr*

En attendant les outils Inserm on-prem:

- Outils NLP souverains de la CE (+ docs.numerique.gouv.fr):

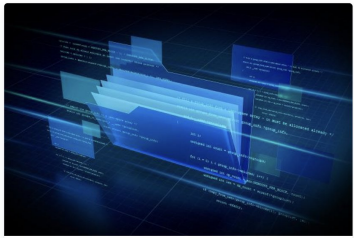
  https://translation.ec.europa.eu/tools-and-resources/ai-translation-and-language-tools_en

All AI-based multilingual tools



eTranslation

The EU institutions' official translation tool is built on decades of translation expertise and offers several styles of translation, including EU formal language. Use safe and secure neural machine translation to translate texts in common formats, or connect your apps and websites for multilingual communication in all EU official languages and some others.

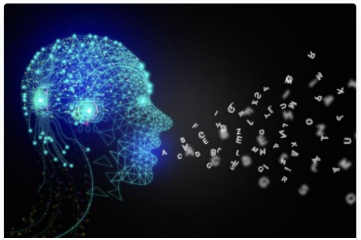


eBriefing

Create background documents from up to 10 source files. Get a first draft in official or general style, with main messages, background and Q&A. Available in all EU official languages and some others.

Anonymisation

Protect privacy by automatically replacing or redacting names, places and other personal data. Use this tool to comply with data protection rules. Available in all EU official languages and some others.



Speech-to-Text

Upload audio or video files for full transcriptions or subtitles. Useful for converting recordings of meetings, workshops or conferences into text. Available in all EU official languages and some others.

Who can use the tools?

- public administrations
- small businesses
- academia
- non-governmental organisations
- Digital Europe programme projects

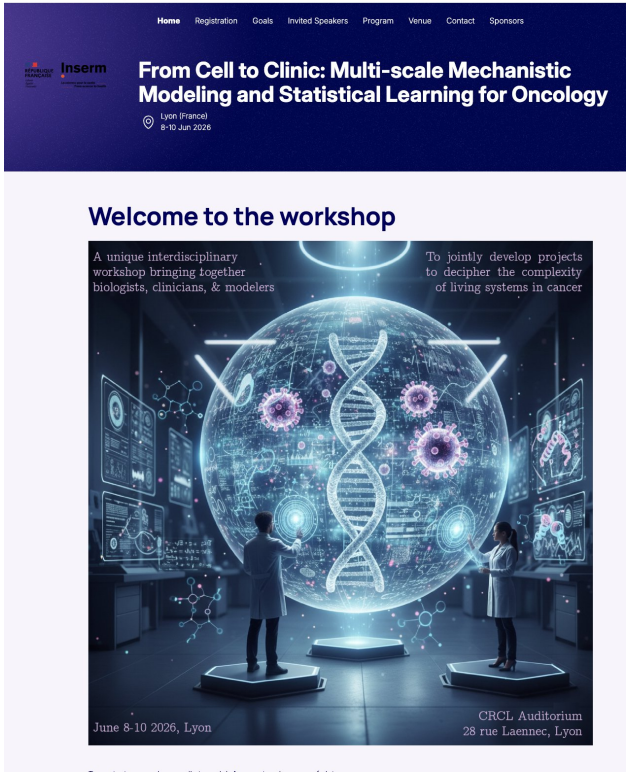
Dynamiser / animer les communautés Inserm IA

Journées scientifiques 2025-2026

AI/ML for the analysis of single-cell spatial transcriptomics, 15-17 Oct 2025, Lyon

AI/ML for Drug Development at Inserm, 23 Jan 2026, Paris (Biopark)

- AI for cellular digital twins in oncology, 8-10 Juin 2026, CLCC, Lyon
- AI for epidemiology, Automne 2026



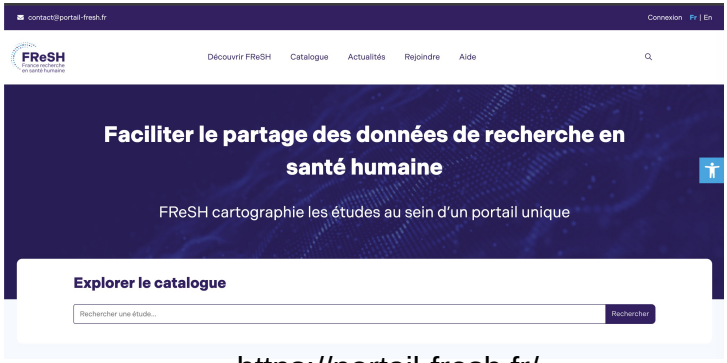
https://models4onco.sciencesconf.org/

Prendre notre part dans les évolutions de l'ecosystème

Evolutions du partage des données

Espace Européen des Donnnées de Santé

- défendre specificités des *données de santé de la recherche* (cohortes, essais, registres)
- pilotage MESR-ONRs-Agences, IT Santé Publique

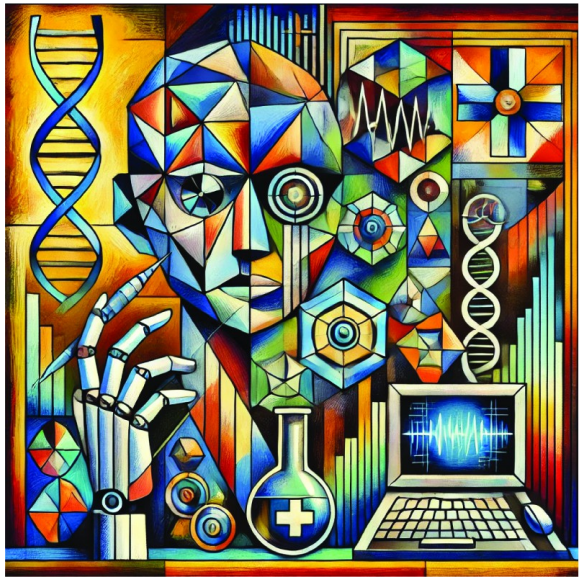


https://portail-fresh.fr/

FReSH comme catalogue

- comité stratégique

MERCI!



généré par chatgpt

Questions?

pole.ia@inserm.fr

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Inserm

La science pour la santé
From science to health