

Exploiter les textes cliniques avec les LLMs : cas d'usage sur le suivi des cancers

Bastien Rance, PhD

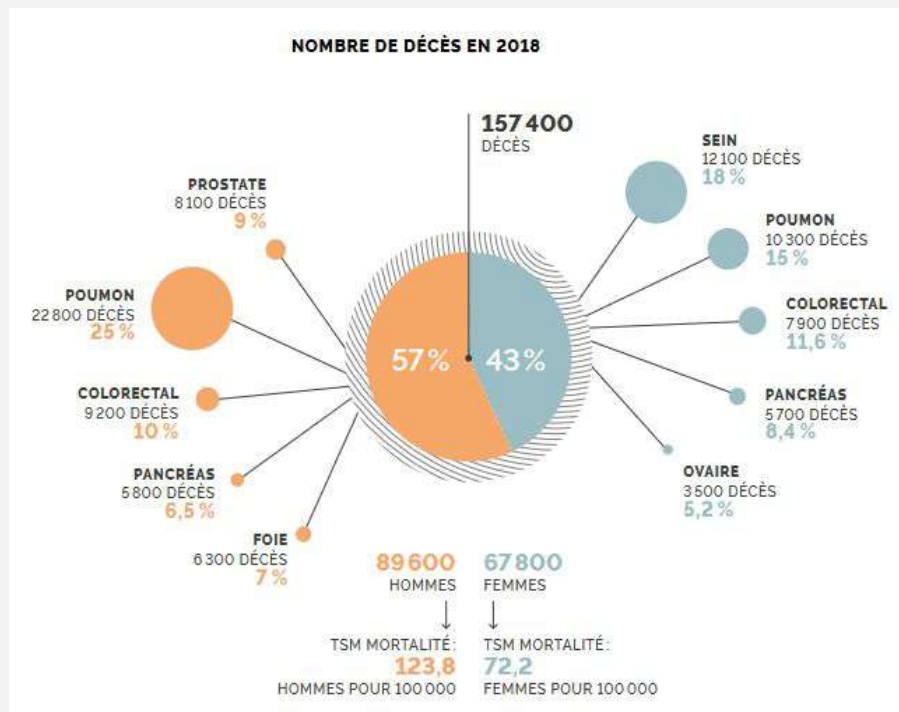
Department of biomedical informatics
HEGP, AP-HP

Inserm, Centre de Recherche des Cordeliers, équipe MUST
(Multi-scale and spatio-temporal modeling in Oncology)





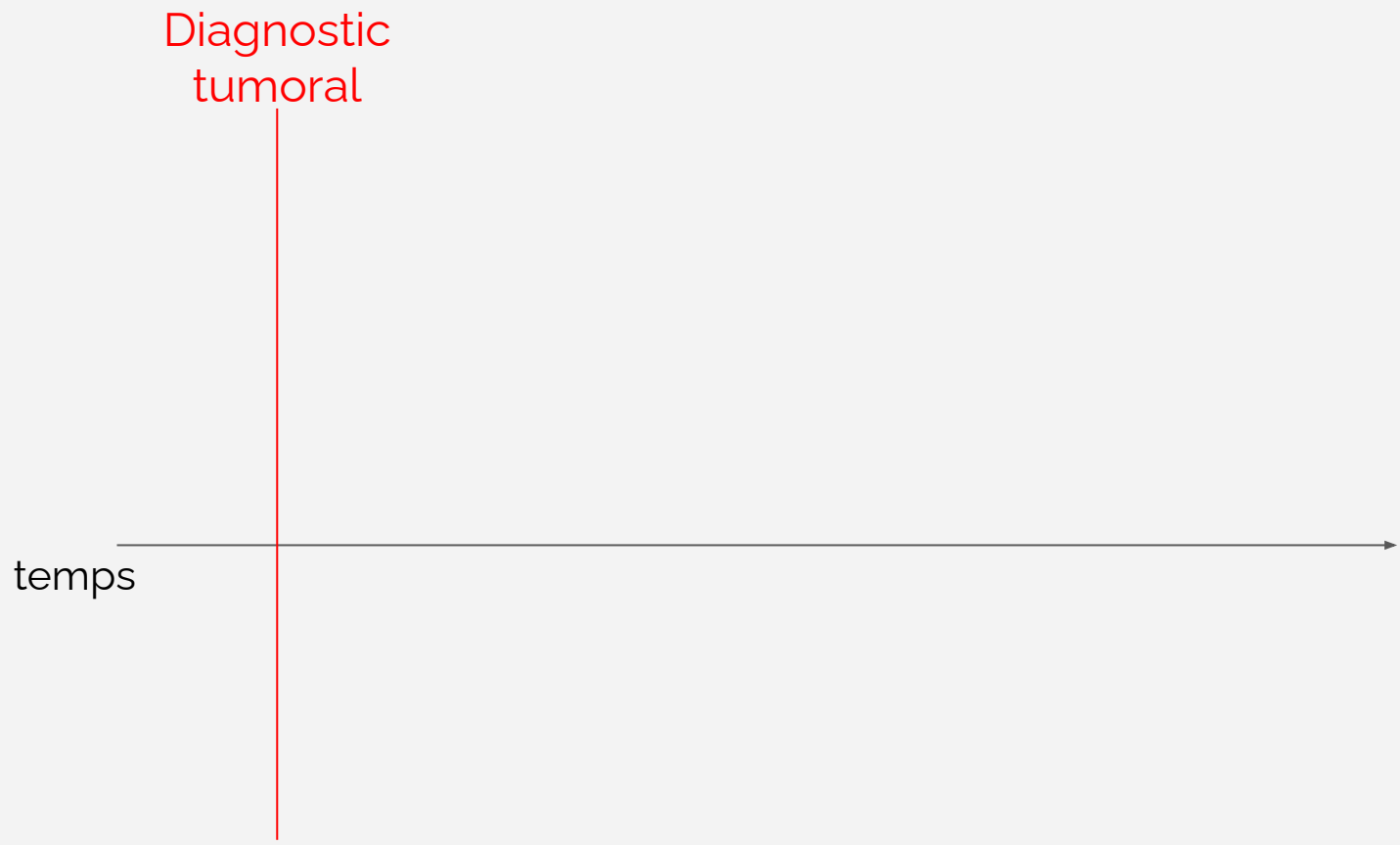


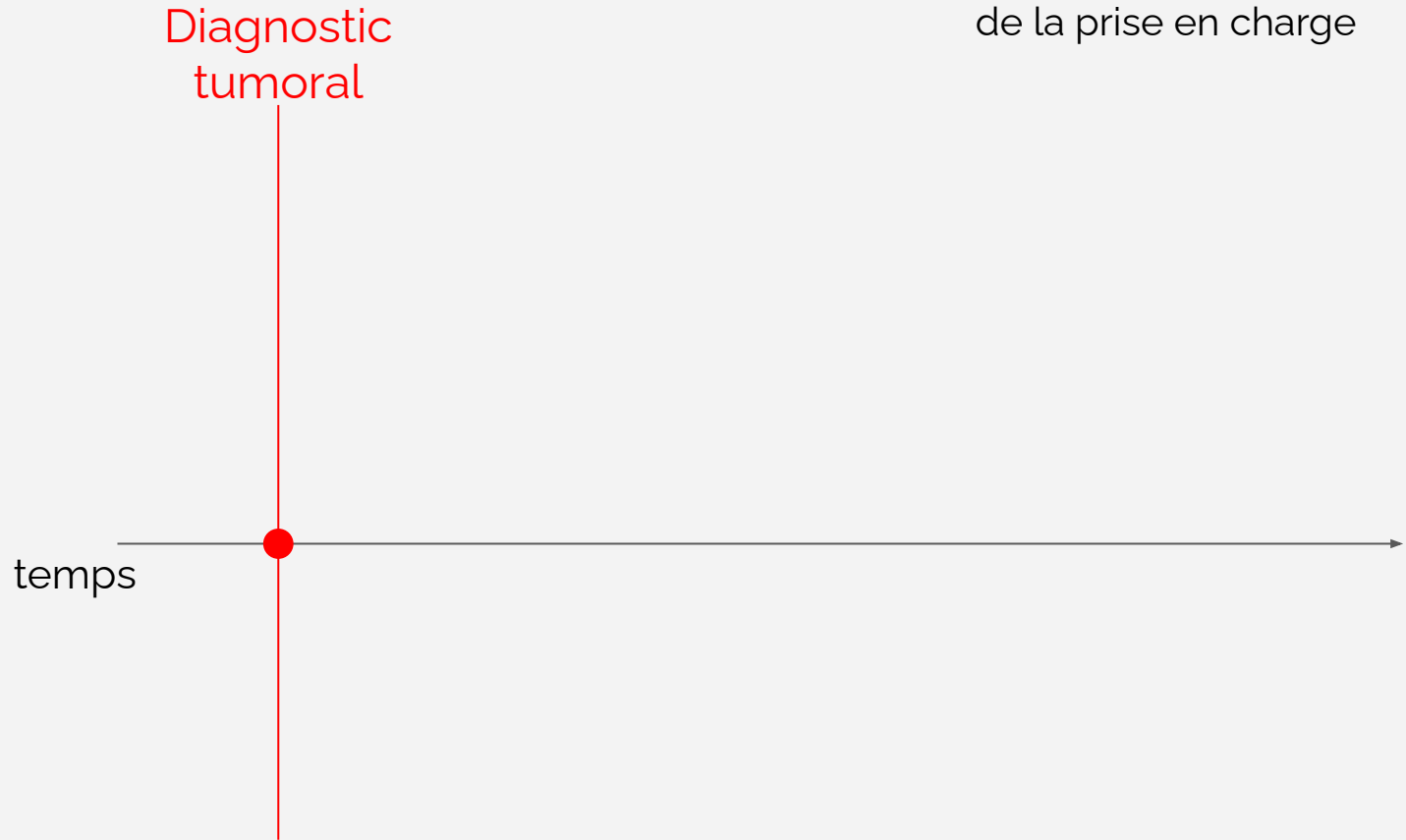


- **1ère** cause de décès chez les hommes (principalement, cancer du poumon)
- **2ème** cause de décès chez la femmes (principalement, cancer du sein)
- Certains cancers en forte augmentation

Source : Santé publique France, en partenariat avec l'Institut national du cancer (INCa) - Epidémiologie et données en cancérologie

Prise en charge et suivi des cancers





Données statiques
collectées au moment
de la prise en charge

Tumor profile

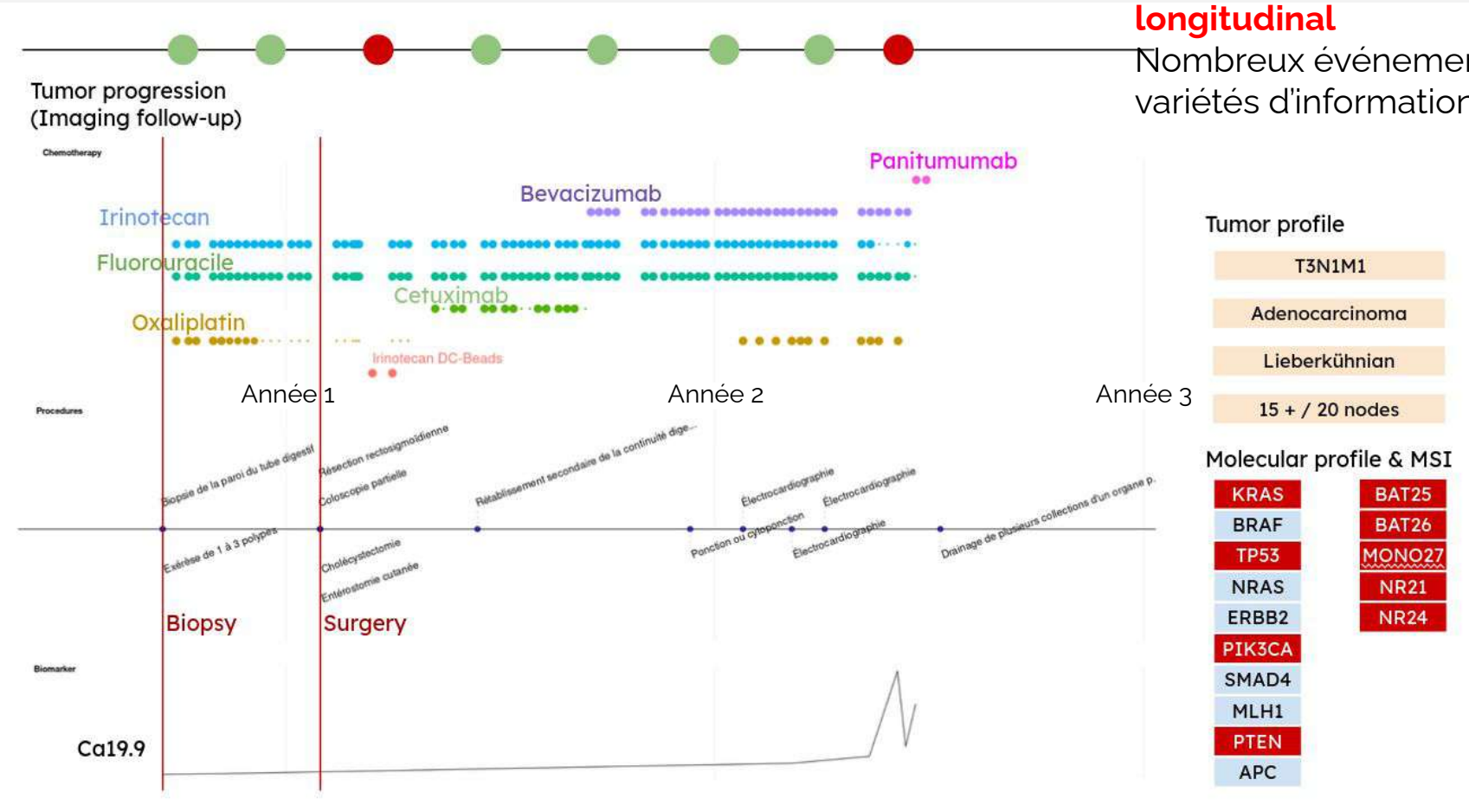
- T3N1M1
- Adenocarcinoma
- Lieberkühnian
- 15 + / 20 nodes

Molecular profile & MSI

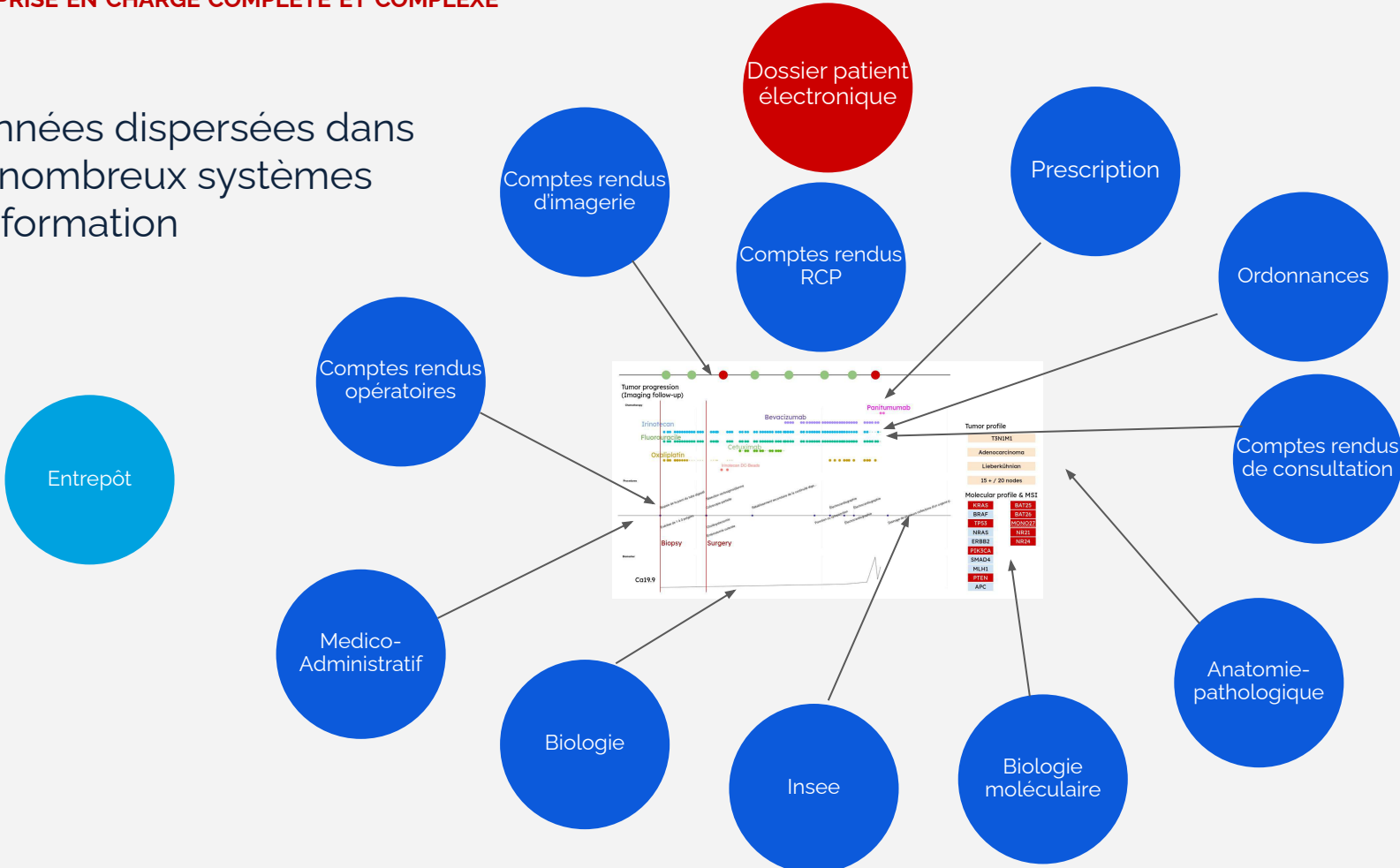
KRAS	BAT25
BRAF	BAT26
TP53	MONO27
NRAS	NR21
ERBB2	NR24
PIK3CA	
SMAD4	
MLH1	
PTEN	
APC	

Traitements et suivi longitudinal

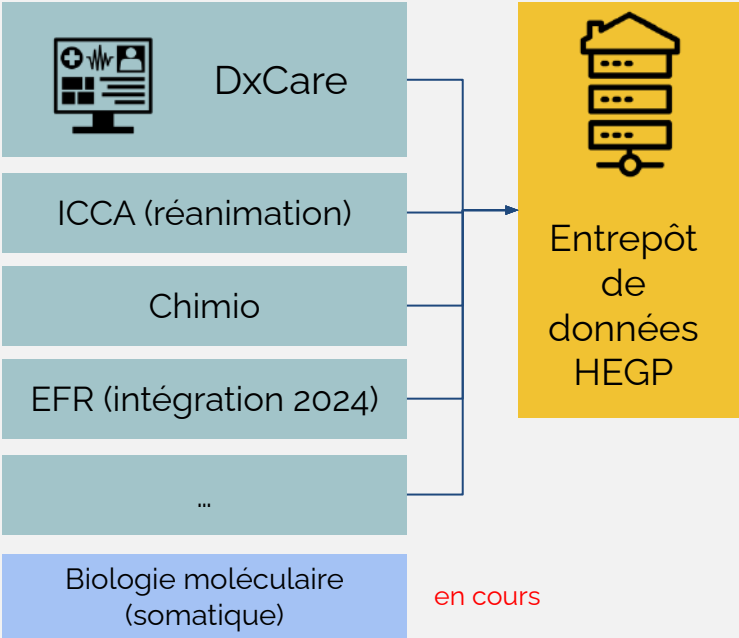
Nombreux événements
variétés d'informations



Données dispersées dans de nombreux systèmes d'information



Entrepôt de données HEGP



Zapletal E, Rodon N, Grabar N, Degoulet P. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. *Stud Health Technol Inform*. **2010**;160(Pt 1):193-7. PMID: 20841676.

Jannot AS, **Zapletal E**, Avillach P, Mamzer MF, **Burgun A**, Degoulet P. The Georges Pompidou University Hospital Clinical Data Warehouse: A 8-years follow-up experience. *Int J Med Inform*. **2017** Jun;102:21-28. doi: 10.1016/j.jimedinf.2017.02.006. Epub 2017 Feb 16. PMID: 28495345.

Zapletal E, Bibault JE, Giraud P, **Burgun A**. Integrating Multimodal Radiation Therapy Data into izb2. *Appl Clin Inform*. 2018 Apr;9(2):377-390. doi: 10.1055/s-0038-1651497. Epub **2018** May 30. PMID: 29847842; PMCID: PMC5976493.

Période 2000 - 2024

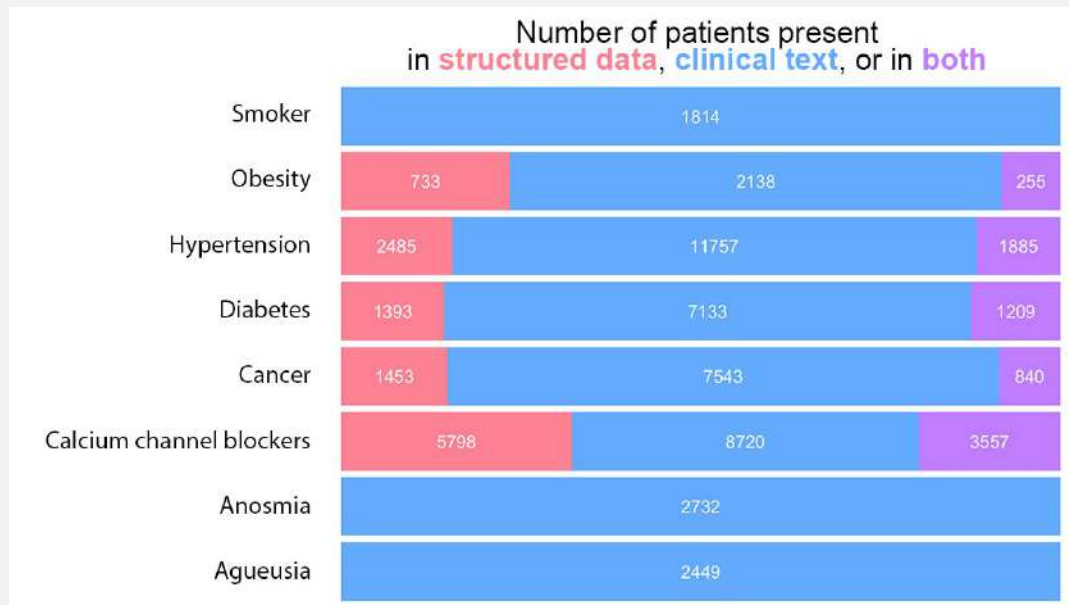
Concept	# observations
Items de questionnaires	311,106,335
Biologie	223,189,999
Données de réanimation	185,756,201
Prescription médicamenteuses	143,032,155
Transmissions infirmières	28,316,055
PMSI	6,640,020
Textes cliniques	7,676,344
Textes complémentaires	2,545,083
Prescription chimio	2,033,379
Temps opératoire	1,674,800

Données non-structurées (textes)

Secondary use of clinical reports - beyond structured data

80%

Of medical information is
present solely in texts



Institut du Cancer Paris

CARPEM

et le projet Master

Observational Trial

En 2024

13 677
patients traités pour un cancer

85 000
séjours en cancérologie

24 500
séances de radiothérapie

34 600
hôpitaux de jour
(en chimiothérapie)

1 286
chirurgies ambulatoires
pour cancer

En chiffres



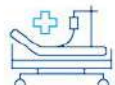
4

HÔPITAUX



6

CENTRES DE RECHERCHE



390

LITS ET 90
FAUTEUILS
D'HÔPITAL DE JOUR



2500

PROFESSIONNELS
DONT 220
MÉDECINS
ET 350 CHERCHEURS



1

COMITÉ DE PATIENTS

HOSPITALITÉ
AP-HP Service Labellisé

34

SECTEURS DE L'INSTITUT
LABELLISÉS



23

PLATEFORMES TECHNIQUES DE POINTE
(Imagerie, radiothérapie, pharmacologie,
biologie spatiale, protéomique...)

CARPEM Master Observational Trial



Données **démographiques**

Données **cliniques**

Données de **génétiques** constitutionnelles
et tumorales cliniques

Données sur les **échantillons**



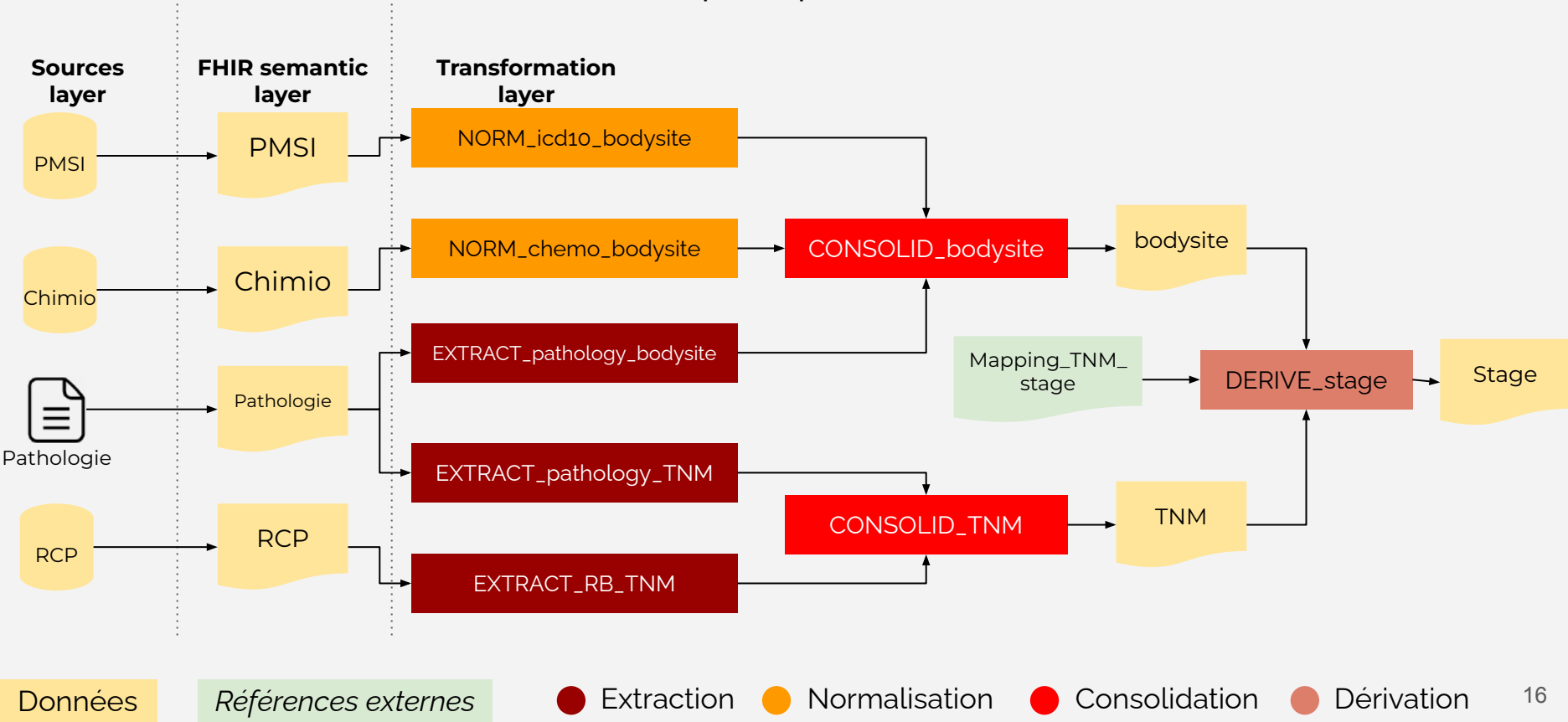
Données **moléculaires**, caractérisation
profonde des échantillons

Données **pathologiques**



AP-HP.Centre-Université Paris Cité

Enjeux: **enrichissement** des données pour permettre la **recherche translationnelle**

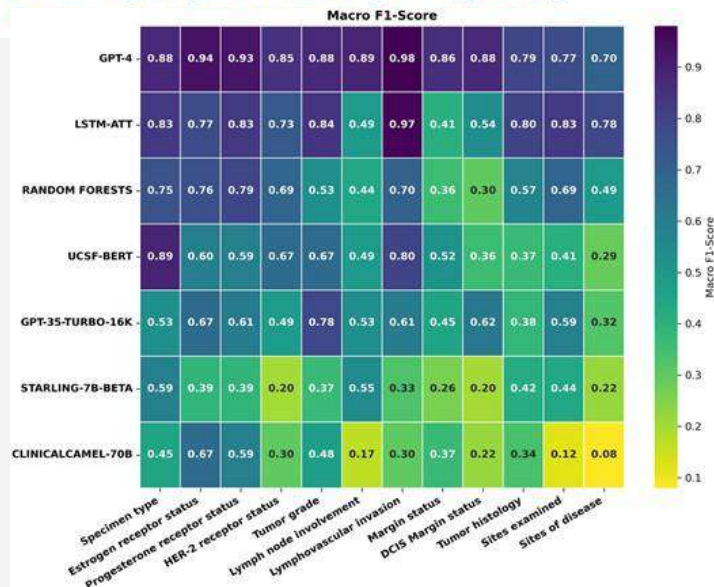


LLMs for data extraction

A comparative study of large language model-based zero-shot inference and task-specific supervised classification of breast cancer pathology reports

Madhumita Sushil, PhD ✉, Travis Zack, MD, PhD, Divneet Mandair, MD,

Zhiwei Zheng, MEng, Ahmed Wali, MEng, Yan-Ning Yu, MEng, Yuwei Quan, MEng,



The test for **estrogen receptors** is positive. There is variable (ranging from weak to strong) nuclear staining in ~70% of tumor cells. Internal positive control is present.

The test for **progesterone receptors** is positive. There is moderate to strong nuclear staining in ~80% of tumor cells. Internal positive control is present.

Result of ***** test: **This carcinoma is negative for ***** oncoprotein over-expression.**

An immunohistochemical assay was performed by manual morphometry on block ***** using the ***** monoclonal antibody to ***** oncoprotein. **The staining intensity of this carcinoma was 1 on a scale of 0-3.** Carcinomas with staining intensity scores of 0 or 1 are considered

FINAL PATHOLOGIC DIAGNOSIS

***** ***** lymph node, left axillary, biopsy:

One lymph node with no tumor identified (0/1). See comment.

B. Left breast, partial mastectomy:

1. Invasive ductal carcinoma, *** grade 2, margins negative for tumor.**

2. Ductal carcinoma in situ, intermediate grade.

3. Non-proliferative fibrocystic change.

See comment.

C. Sentinel lymph node, left axillary, biopsy:

One lymph node with no tumor identified (0/1).

Site examined: ☒ Left Breast¹ ☒ Left LN¹ ☐ Right Breast¹ ☐ Right LN¹ ☐ Other tissues¹ ☐ Unknown¹

Sites of disease: ☒ Left Breast¹ ☐ Left LN¹ ☐ Right Breast¹ ☐ Right LN¹ ☐ Other tissues¹ ☐ None¹ ☐ Unknown¹

Histology: ☐ No malignancy¹ ☐ LCIS¹ ☒ DCIS ☒ Invasive ductal ☐ Invasive lobular ☐ Medullary ☐ Mucinous ☐ Tubular ☐ Papillary ☐ Metaplastic BC ☐ Mixed ☐ Carcinoma NOS ☐ Unknown

LN involvement: ☒ 0 involved ☐ 1-3 involved ☐ 4-9 involved ☐ 10+ involved ☐ Unknown

Biopsy type: ☐ Biopsy ☐ Lumpectomy ☒ Mastectomy ☐ Unknown

ER: ☐ Low positive ☒ Positive ☐ Negative ☐ Unknown

PR: ☒ Positive ☐ Negative ☐ Unknown

HER2: ☐ Positive ☒ Negative ☐ Equivocal ☐ Equivocal Positive ☐ Equivocal Negative ☐ Unknown

Max grade: ☐ 0 ☐ 1 (Low) ☒ 2 (Intermediate) ☐ 3 (High) ☐ Unknown

LVI: ☒ Present ☐ Absent ☐ Unknown

Margins: ☐ Positive margin ☒ Less than 2mm ☐ More than/eq to 2mm ☐ Unknown

DCIS Margins: ☐ Positive margin ☒ Less than 2mm ☐ More than/eq to 2mm ☐ Unknown

"Across all 13 tasks, the **GPT-4 model** performed either significantly better than or as well as the best supervised model, the LSTM-Att model"

JOURNAL ARTICLE FEATURED

Evaluating the effectiveness of biomedical fine-tuning for large language models on clinical tasks

[Get access >](#)

Felix J Dorfner, Amin Dada, MSc, Felix Busch, MD, Marcus R Makowski, MD, PhD, Tianyu Han, PhD, Daniel Truhn, MD, MSc, Jens Kleesiek, MD, PhD, Madhumita Sushil, PhD, Lisa C Adams, MD ✉, Keno K Bressemer, MD

Journal of the American Medical Informatics Association, Volume 32, Issue 6, June 2025, Pages 1015–1024, <https://doi.org/10.1093/jamia/ocaf045>

Published: 07 April 2025 **Article history** ▼



“Biomedical LLMs **generally underperformed** compared to **general-purpose models**, especially on tasks not focused on probing medical knowledge”

Felix J Dorfner, Amin Dada, Felix Busch, Marcus R Makowski, Tianyu Han, Daniel Truhn, Jens Kleesiek, Madhumita Sushil, Lisa C Adams, Keno K Bressemer, Evaluating the effectiveness of biomedical fine-tuning for large language models on clinical tasks, *Journal of the American Medical Informatics Association*, Volume 32, Issue 6, June 2025, Pages 1015–1024, <https://doi.org/10.1093/jamia/ocaf045>

Données collectées sous le régime de la **non-opposition**

→ **Impossible d'utiliser** des **LLM commerciaux**

(Nécessiterait le consentement des patients, coût difficile à contrôler)

Ressources limitées dans l'hôpital

→ Serveur modéré 4x A40 48GB + 1x H100 96 GB

Peu d'accès à de l'**annotation experte**

→ Temps médical précieux, difficulté à rémunérer des experts

Intérêt du fine-tuning pour des tâches spécifiques ?

Amélioration des performances via instruction-tuning ?

Exploration des hyper-paramètres et des stratégies d'inférence des LLMs

Et comment optimiser le LLM pour cet usage ?

Examen réalisé dans le cadre de la **surveillance d'un adénocarcinome colique** connu, traité par chirurgie puis chimiothérapie. Le patient se présente pour un contrôle d'imagerie afin d'évaluer l'**évolution des lésions** déjà identifiées lors du précédent examen.

On retrouve au niveau de la loge colique droite l'aspect post-opératoire habituel, **sans collection ni modification notable**. Le **site tumoral** initialement décrit comme stable lors du dernier contrôle **ne montre pas de nouvelle anomalie focale** évidente. Aucun épaissement pariétal suspect n'est visible sur les segments coliques explorés, et le rehaussement mural reste homogène.

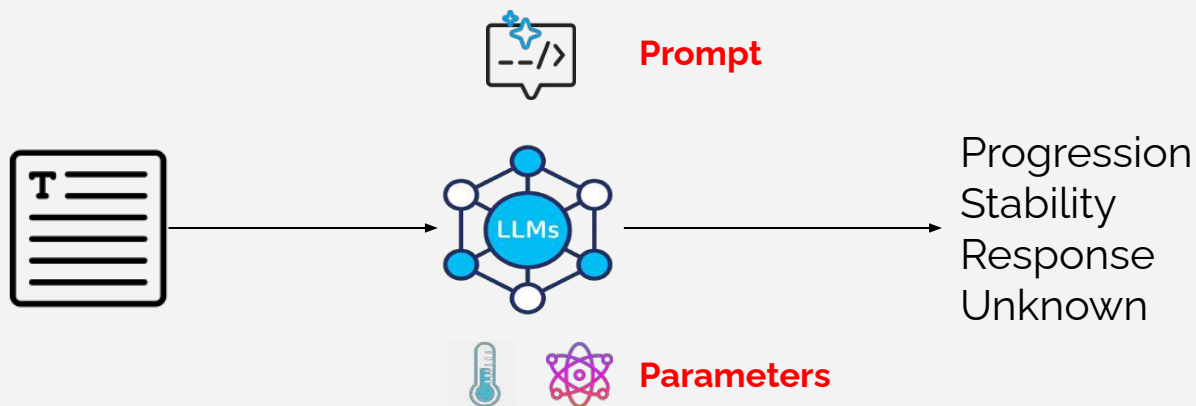
Au niveau hépatique, **les deux lésions segmentaires précédemment mesurées sont à nouveau identifiées**. La lésion du segment VI, mesurée antérieurement à 18 mm dans son plus grand axe, apparaît aujourd'hui à 20 mm, **soit une légère augmentation dimensionnelle**. La lésion du segment III, jusque-là de 12 mm, se maintient à 12-13 mm selon le plan de coupe. Les contours demeurent nets, sans modification du caractère de rehaussement. **L'ensemble est compatible avec une progression faible** mais réelle selon les critères RECIST, principalement liée à l'augmentation de la lésion segment VI, même si la variation reste à la limite du seuil interprétatif. Il n'apparaît **pas de nouvelle lésion hépatique**.

Les chaînes ganglionnaires abdomino-pelviennes ne montrent **pas d'adénopathie dépassant le seuil RECIST** habituel (10 mm en court axe). **Quelques ganglions mésentériques** persistants, de petite taille, restent **identiques aux examens antérieurs**. Pas d'épanchement péritonéal observé. Le parenchyme pulmonaire visualisé à la limite du champ d'acquisition ne révèle pas d'anomalie suspecte.

Can LLM understand radiology reports?

Experience : Exploration of Hyperparameters Configurations and their Impact on Inference Stability.

Valentin Pohyer
In preparation



Example of a prompt:

INSTRUCTIONS

You are an expert French radiologist. You are reading reports of imaging studies (for example CT or MRI or PET-CT) that have been performed to evaluate the tumor's response to treatment. Your goal is to determine whether the cancer is responding to treatment, is stable or is progressing. There can be imaging reports where the evolution of the cancer is unknown or studies which are not a follow-up. Your task is to choose from one of five following classes according to the radiology report: Progression, Stability, Response, Unknown or Not-a-follow-up.

The output has to be a json object.

Here is an example of the expected output format : `{{"cancer_evolution": "Progression"}}`

You will only respond with the json code. Do not provide explanations or notes. Do not generate empty lines.

EXAMPLES

Text input : Au total : SD - PD - Oui : PD

Json output : `{{"cancer_evolution": "Progression"}}`

Text input : Maladie progressive

Json output : `{{"cancer_evolution": "Progression"}}`

Text input : Aspect TEP d'évolutivité néoplasique

Json output : `{{"cancer_evolution": "Progression"}}`

Text input : Au total : SD + NA + non = SD (soit maladie stable)

Json output : `{{"cancer_evolution": "Stability"}}`

| ...]

Text input : Bilan initial

Json output : `{{"cancer_evolution": "Not-a-follow-up"}}`

Text input : Examen de réactualisation avant chirurgie

Json output : `{{"cancer_evolution": "Not-a-follow-up"}}`

IMAGING REPORT

[[TEXT REPORT]]

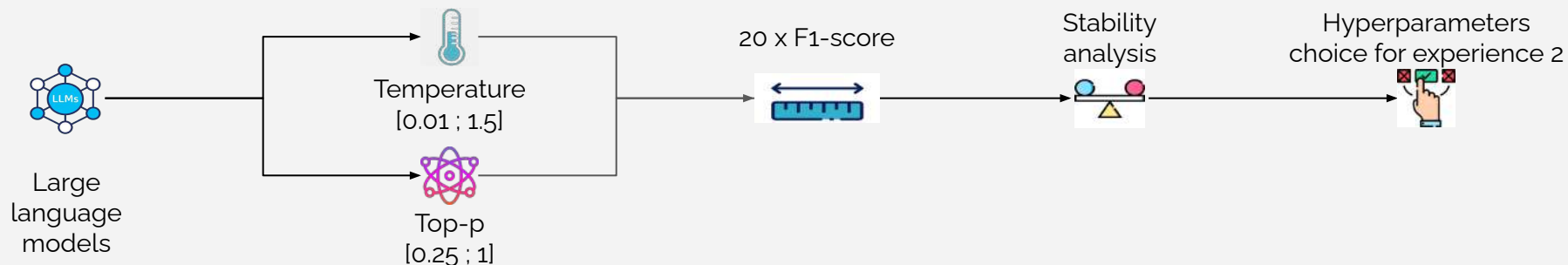
OUTPUT

`{{"cancer_evolution": "`

LARGE LANGUAGE MODELS - OPPORTUNITIES AND CHALLENGES

Experience 1 : Exploration of Hyperparameters Configurations and their Impact on Inference Stability.

Valentin Pohyer
In preparation



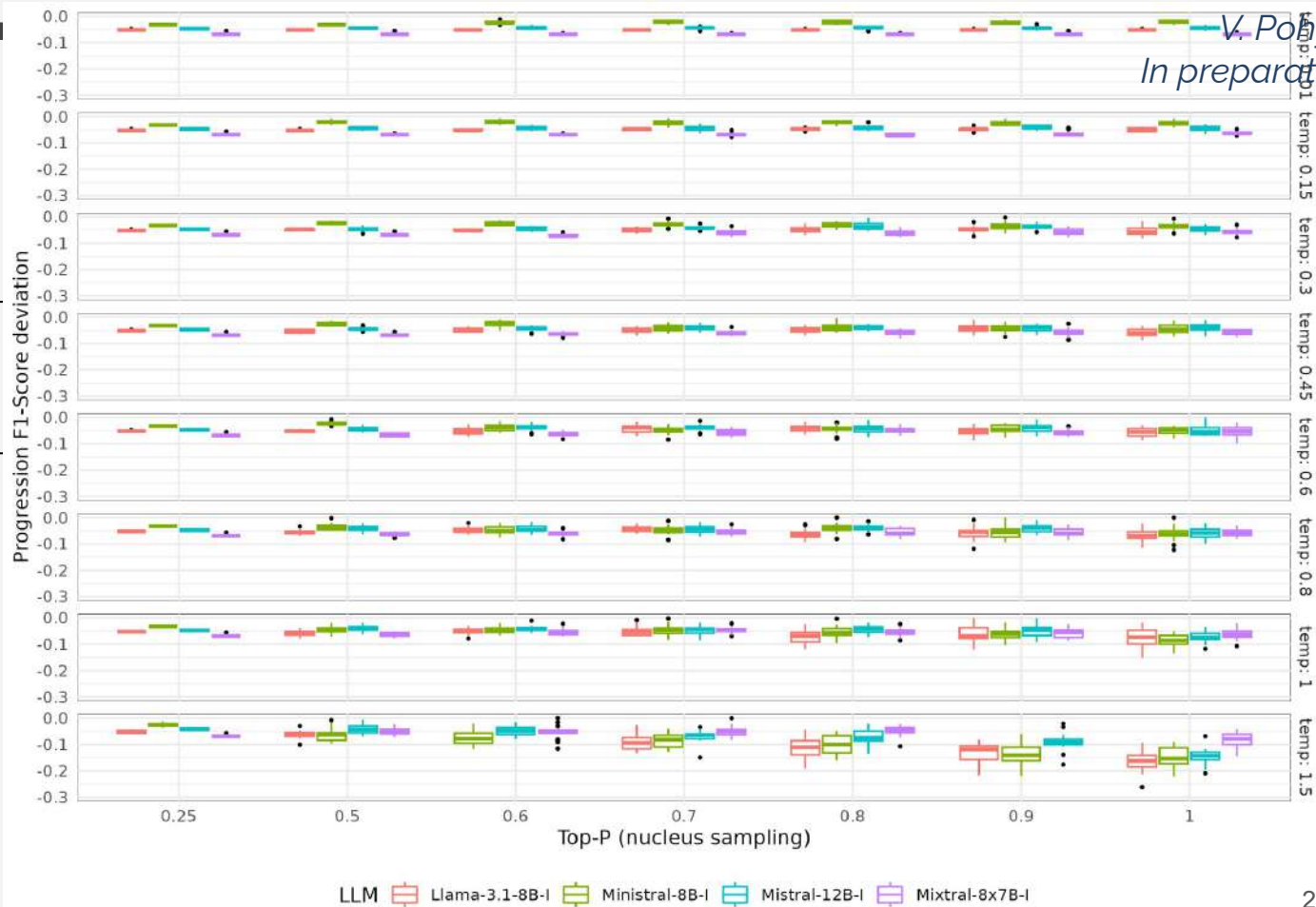
The diagram shows a comparison of model outputs. On the left, a blue downward arrow and a thermometer icon point to the word 'jour' in large purple font. Below it, the words 'monsieur' and 'matin' are written in smaller purple font. On the right, a red upward arrow and a thermometer icon point to the word 'jour' in large purple font. Below it, the words 'monsieur' and 'matin' are written in smaller purple font. A small thermometer icon is also present to the left of the 'jour' text on the left side.

LARGE LANGUAGE MODELS - OPPORTUNITIES AND CHALLENGES

Experience 1 : Exploration of Hyper



Large
language
models



V. Pothier
In preparation

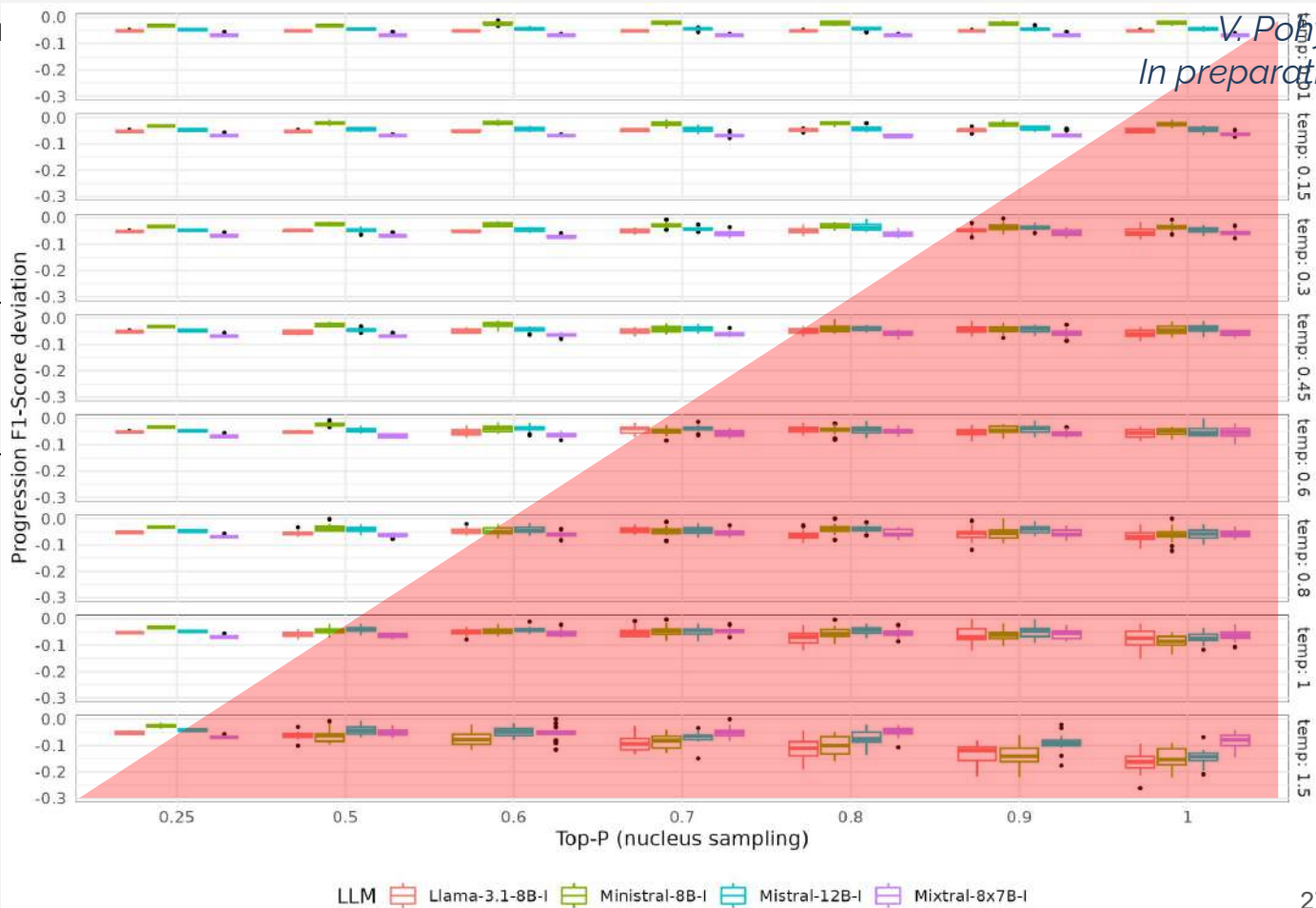
LARGE LANGUAGE MODELS - OPPORTUNITIES AND CHALLENGES

Experience 1 : Exploration of Hyper



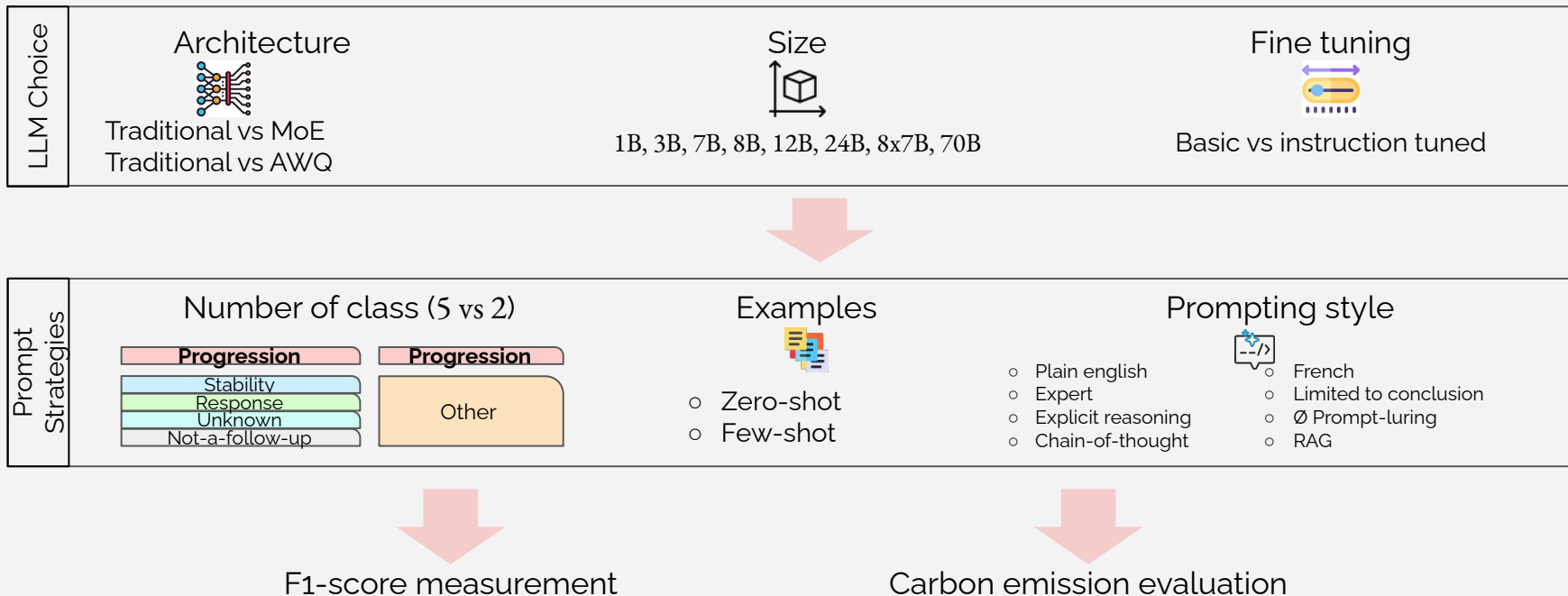
Large
language
models

→ Limiter les top-p et
la température



Experience 2 : Exploration of Architecture, Size, Fine-tuning and Prompting Strategies.

Valentin Pohyer
In preparation



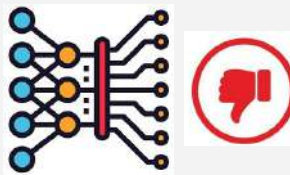
RÉSULTATS

Valentin Pohyer
In preparation

Traditionnel Mixture-of-Experts

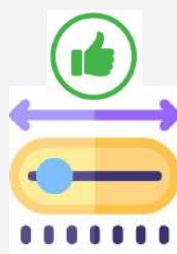


Mistral-**12B**
0.78 [0.77 - 0.80]*



Mixtral-**8x7B**
0.72 [0.70 - 0.74]*

Fine-tuning instruction ($p = 1.43 \times 10^{-6}$)



Llama-1B	0.35 [0.34 - 0.36]*
Llama-1B- instruct	0.44 [0.42 - 0.45]*
	+0.09
Llama-8B	0.51 [0.50 - 0.53]*
Llama-8B- instruct	0.70 [0.68 - 0.72]*
	+0.19

Taille



Llama-**70B**
0.77 [0.75 - 0.79]*



Llama-**8B**
0.70 [0.68 - 0.72]*



Llama-**3B**
0.55 [0.53 - 0.57]*



Llama-**1B**
0.44 [0.42 - 0.45]*

*prompt :
Plain English
5 classes
Few-Shot

RÉSULTATS

Valentin Pohyer
In preparation

Nombre de classe (5 VS 2)

$$p = 4.26 \times 10^{-8}$$

5 classes
0.616 en
moyenne



Progression
Stabilité
Réponse
Inconnue
Pas-un-suivi

Progression
Autre

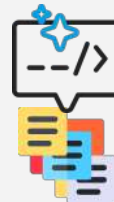
2 classes
0.592 en
moyenne



Exemples

$$p = 5.86 \times 10^{-6}$$

Few-shot
0.624 en
moyenne



Zero-shot
0.601 en
Moyenne



Expertise médicale



Expert
0.73 [0.70 - 0.77]*



Plain English
0.66 [0.64 - 0.68]*



Prompt luring



Avec
0.70 [0.68 - 0.72]*

30



Sans
0.29 [0.24 - 0.35]*

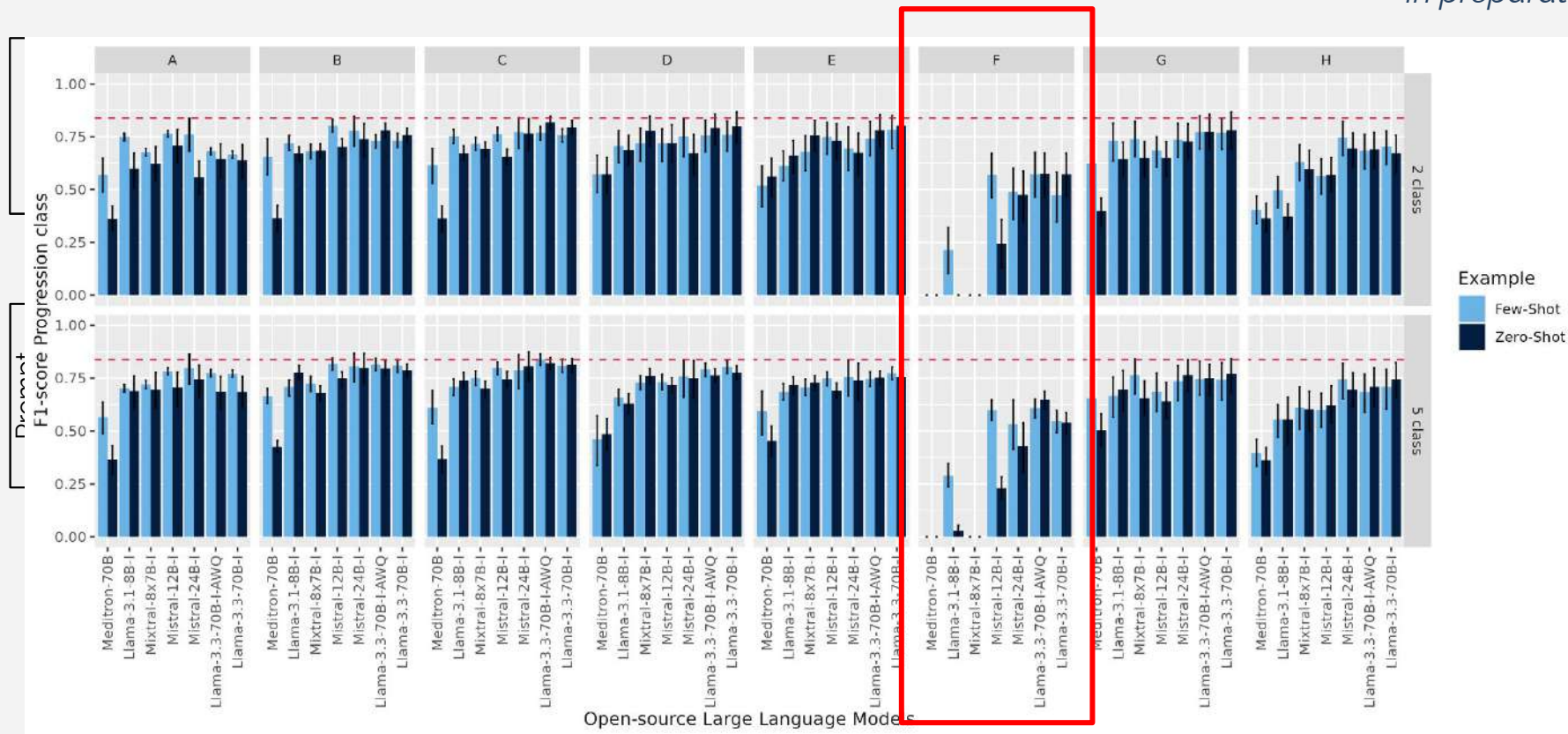


*prompt :
Plain English
5 Classes
Few-Shot

LARGE LANGUAGE MODELS - OPPORTUNITIES AND CHALLENGES

Experience 2 : Exploration of Architecture, Size, Fine-tuning and Prompting Strategies.

Valentin Pohyer
In preparation

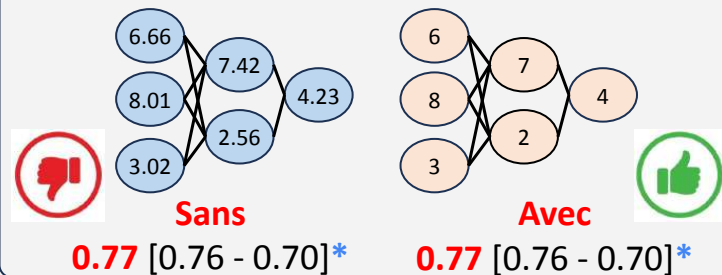


No prompt luring

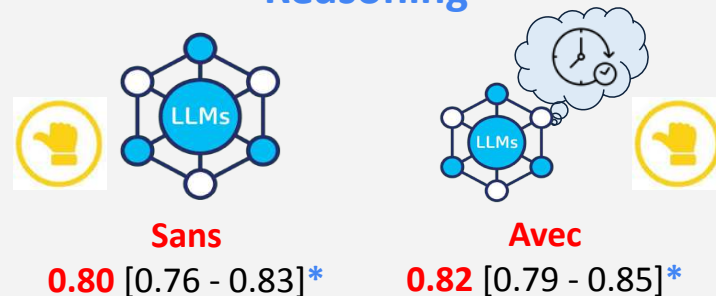
RÉSULTATS

Valentin Pohyer
In preparation

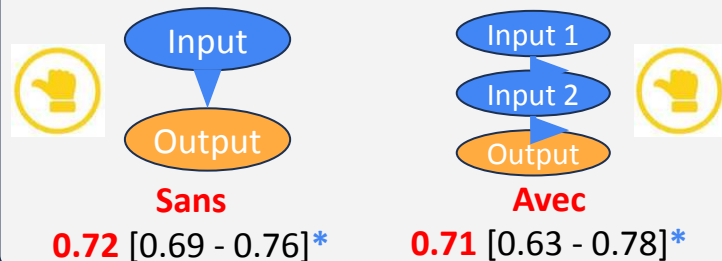
Quantization



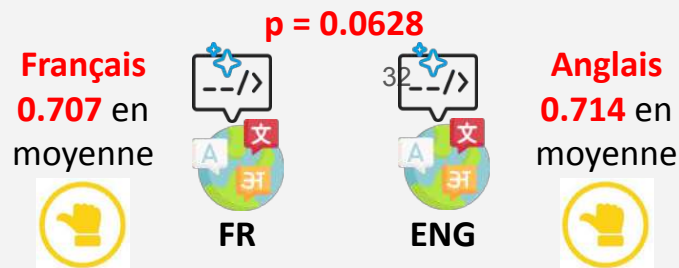
Reasoning



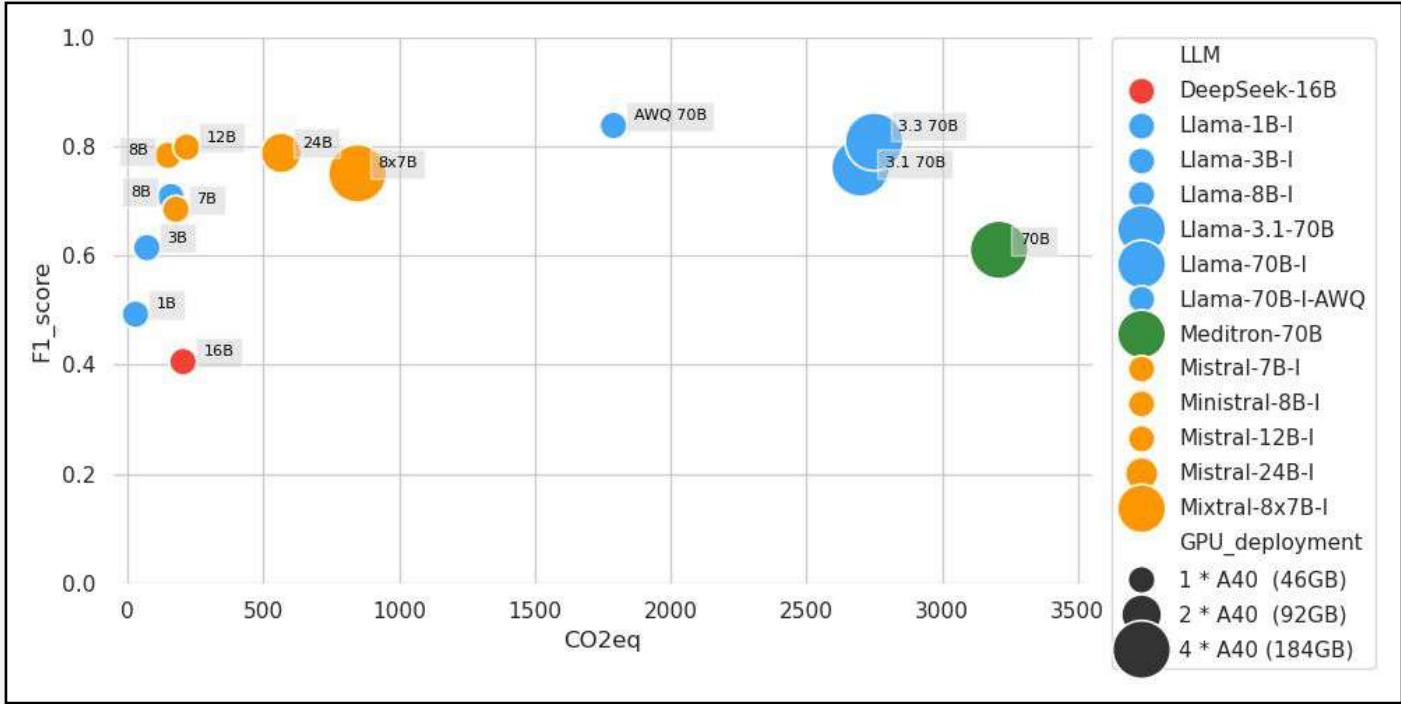
Chain-Of-Thought



Langue utilisée



*prompt :
Plain English
5 Classes
Few-Shot



communications medicine

[Explore content](#) [About the journal](#) [Publish with us](#)

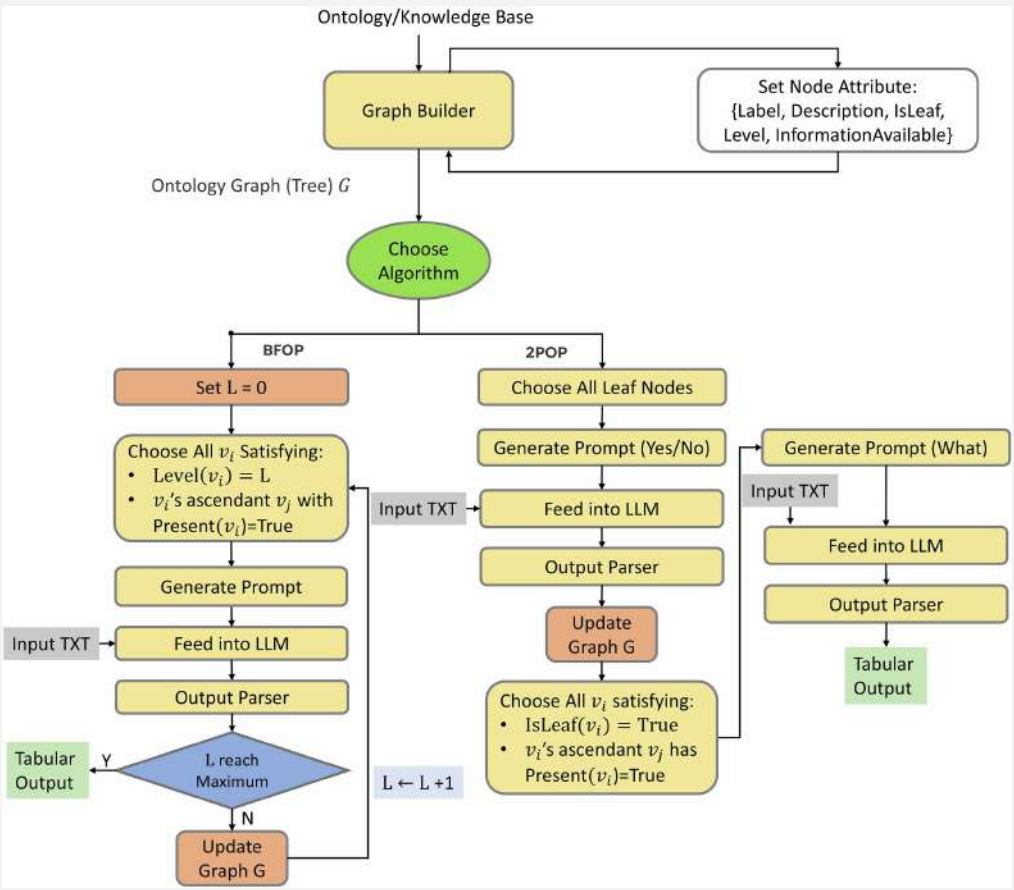
[nature](#) > [communications medicine](#) > [articles](#) > [article](#)

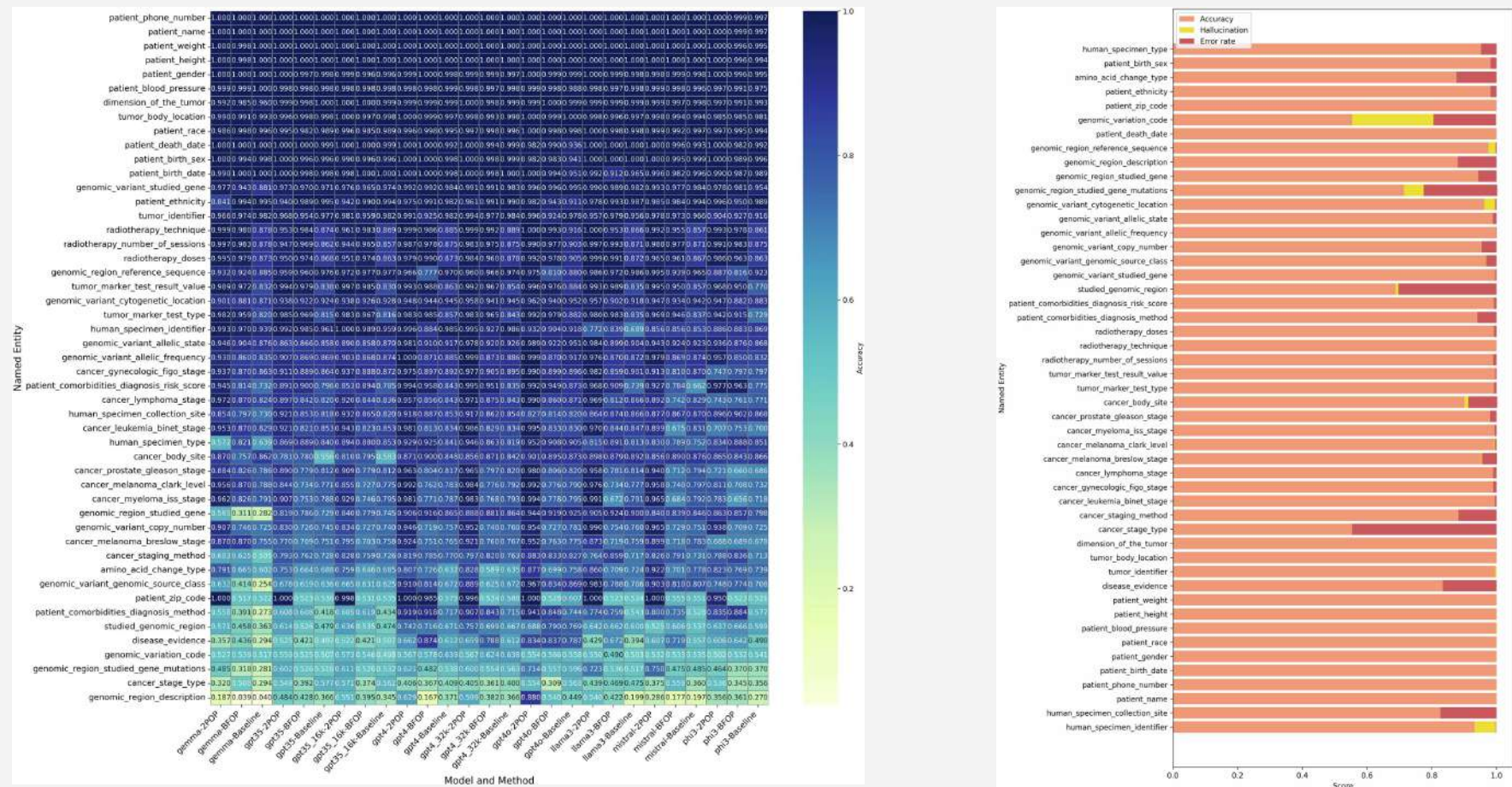
Article | [Open access](#) | Published: 15 October 2025

Introducing mCODEGPT as a zero-shot information extraction from clinical free text data tool for cancer research

[Kai Zhang](#), [Tongtong Huang](#), [Bradley A. Malin](#), [Travis Osterman](#), [Qi Long](#) & [Xiaoqian Jiang](#) 

[Communications Medicine](#) **5**, Article number: 422 (2025) | [Cite this article](#)



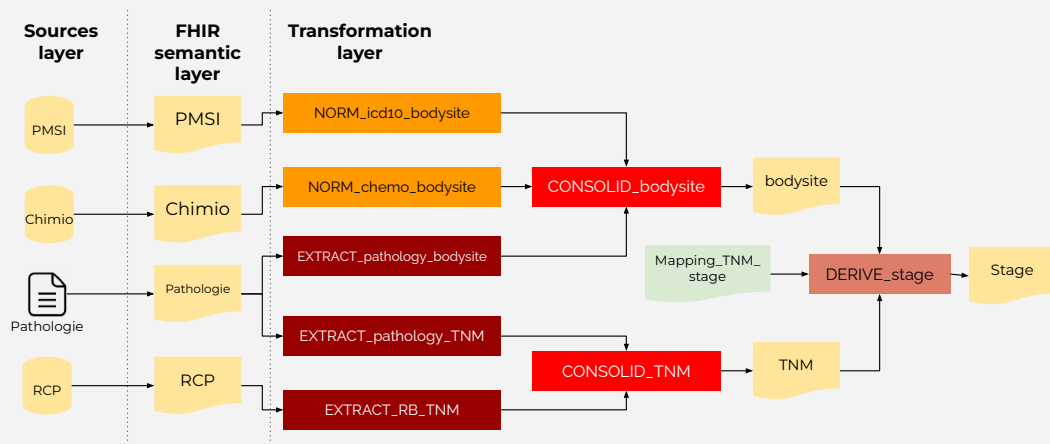


Zhang, K., Huang, T., Malin, B.A. *et al.* Introducing mCODEGPT as a zero-shot information extraction from clinical free text data tool for cancer research. *Commun Med* **5**, 422 (2025). <https://doi.org/10.1038/s43856-025-01116-x>

Gouverner les extractions

Développement de nombreux **extracteurs**, e.g.:

- Expressions régulières
- Transformers
- LLMs
 - Avec ou sans RAG

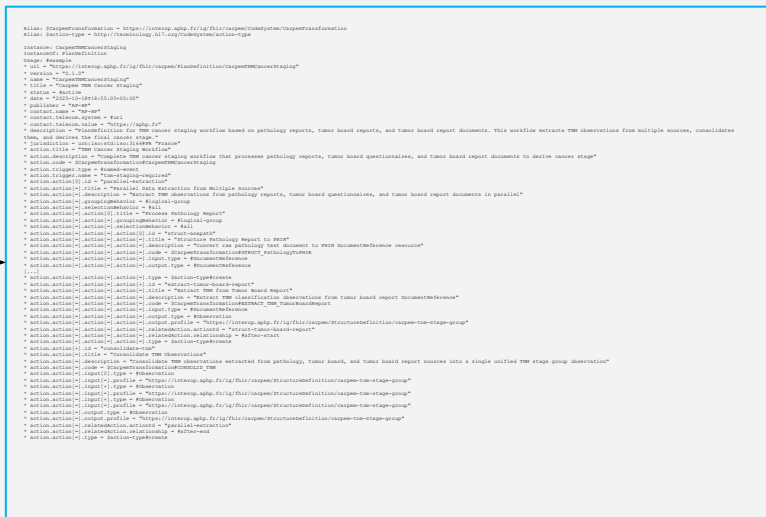


Workflow d'extraction complexe

- Gestion des inputs et output
 - Alignement avec des standards d'interopérabilité (HL7 FHIR)
- Gestion du DAG des processus d'extraction

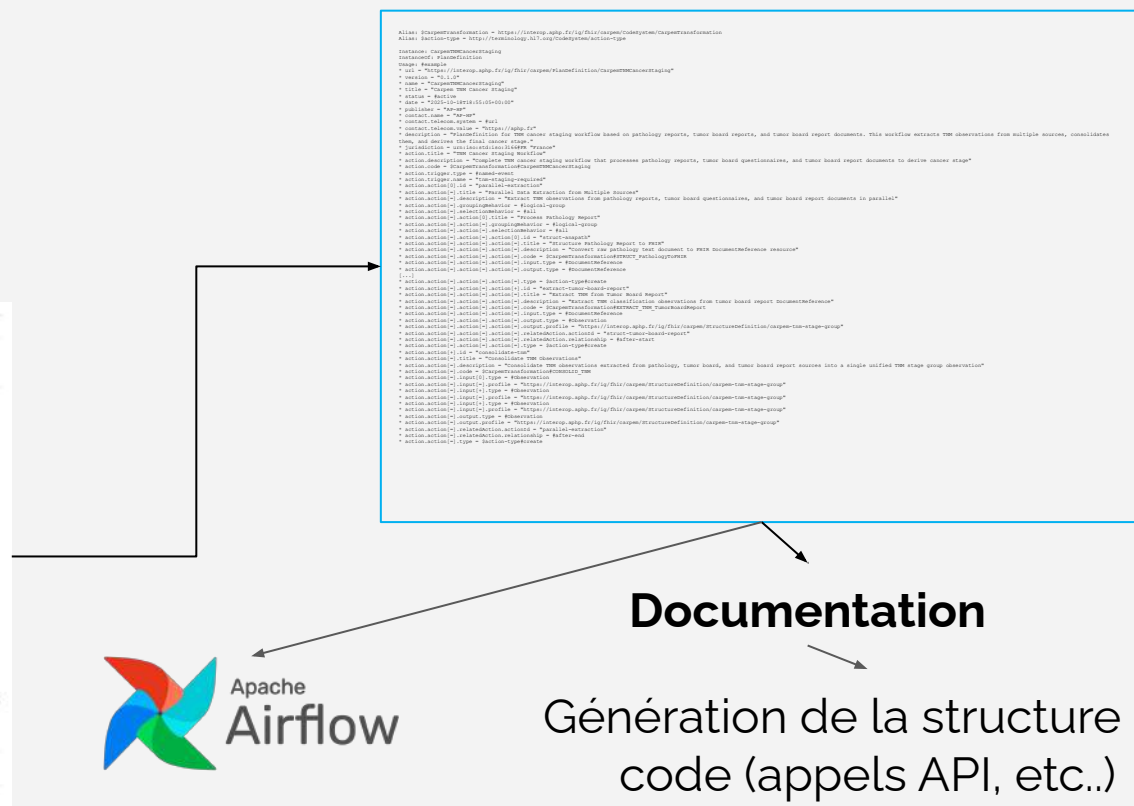
Toward documentation-as-code with LLMs ?

Génération d'un artefact **FHIR** PlanDefinition (json)



Documentation

Génération d'un artefact **FHIR** PlanDefinition (json)



Documentation

Génération de la structure du code (appels API, etc..)

Ce que les LLMs ne savent pas (encore) faire

- Résoudre les **contradictions entre sources**
 - → Mais aide à la documentation
- Intégrer le **temps long du suivi longitudinal** multi-CR
 - ... **Problème ouvert** sur les textes médicaux
- Garantir une **reproductibilité** stricte
 - → Fort besoin de **traçabilité**

- LLMs performants mais **non déterministes** → **contrôle hyperparamètres indispensable**.
- **Instruction-tuning + prompting expert** > taille du modèle.
- Contexte **hospitalier** = contraintes **structurantes** (local, peu de GPU, pas de cloud).
- Le vrai défi est la **gouvernance** du pipeline (FHIR, documentation-as-code).
- Vers un **NLP clinique fiable, stable et auditable**.

The Team

PhD students

Valentin Pohyer (PhD ongoing)

Engineers

Hector Countouris

David Baudoin

Joséphine Bocquet

Mehdi Ismaïli

Vincent Zossou

Thank you!

CARPEM Master Observational Trial

Core team

Sandrine Pham

Bastien Rance

Vincent Zossou

Ethics and regulatory advisor

Marie-France Mamzer

Advisor data governance and interoperability

David Ouagne

**In collaboration with all the clinicians and
researchers of the CARPEM**

