

# DEFT 2023

## Défi Fouille de Textes@TALN 2023



**Organisé par :**

Laboratoire Informatique d'Avignon (LIA)  
Laboratoire des Sciences du Numérique de Nantes (LS2N)



**Lundi 5 mai 2023**

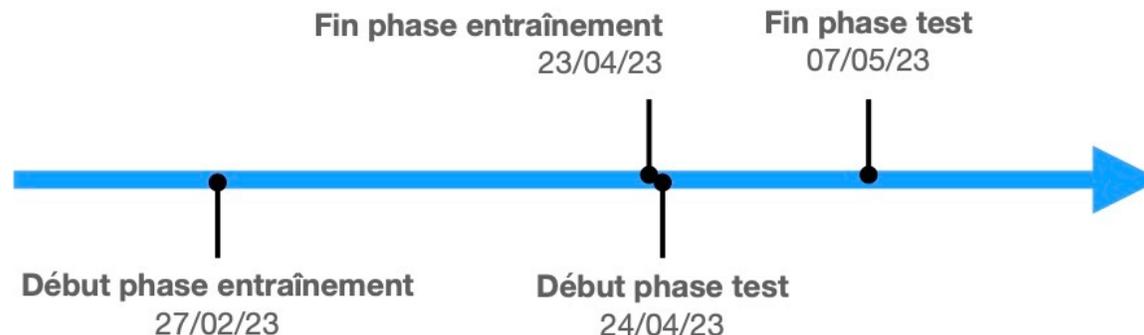
Sorbonne Center for Artificial Intelligence (SCAI)  
Paris, France

<http://deft2023.univ-avignon.fr>



# Contexte

- Campagne d'évaluation annuelle francophone
- Tâches de traitement automatique du langage (TAL)
- Confrontation d'approches variées
- Lieu d'échange pour les acteurs du TAL
  - Milieu académique / industriel
  - Différents niveaux d'expertise (étudiants, débutants, confirmés...)

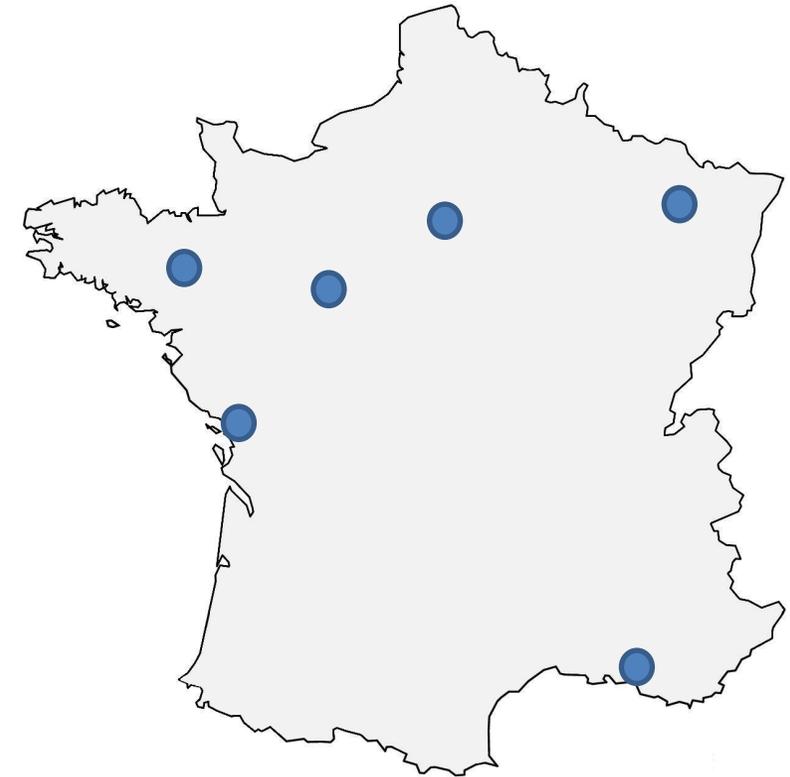


# Objectifs et tâches de l'atelier

- Type de données : Questions à choix multiples (QCMs) en français issus d'annales d'examens de pharmacie
- Difficultés :
  - Domaine de spécialité
  - Plusieurs réponses possibles par question
- Deux tâches proposées :
  - Principale : Sélectionner automatiquement la ou les réponse(s) correcte(s)
  - Annexe : Identifier le nombre exact de réponses correctes
- Deux pistes de recherche :
  - Reproductible : Données et modèles connus (classement officiel)
  - Aucune restriction : Utilisation de tout modèle possible (pas de classement)

# Équipes

- Six équipes participantes (ordre alphabétique) :
  - ALMANACH-ARKHN : Équipe jointe entre l'entreprise Arkhn et l'INRIA
  - LIS : Aix-Marseille Université
  - LIUM-IRISA : Équipe jointe entre le LIUM (Le Mans Université) et l'IRISA (Université de Rennes)
  - SEQUOIA : Entreprise EDF R&D
  - SPQR : Équipe jointe entre le STIH (Sorbonne Université), le L3I (La Rochelle Université) et l'OBTIC (Sorbonne Université)
  - TTGV : Équipe jointe entre l'entreprise SNCF R&D, le LORIA (Université de Lorraine) et le LTCl (Telecom Paris)



# Données

- Corpus FrenchMedMCQA de QCMs
  - Par question : 5 réponses potentielles + réponse(s) correcte(s) identifiée(s)
  - Annales d'examens réels de pharmacie
  - Obtenu à partir du site *Remede.org*
  - Manuellement réalisé par des experts médicaux

# Réponses	Apprentissage	Développement	Évaluation	Total
<b>1</b>	595	164	321	<b>1 080</b>
<b>2</b>	528	45	97	<b>670</b>
<b>3</b>	718	71	141	<b>930</b>
<b>4</b>	296	30	56	<b>382</b>
<b>5</b>	34	2	7	<b>43</b>
<b>Total</b>	<b>2 171</b>	<b>312</b>	<b>622</b>	<b>3 105</b>

# Exemple d'instance

```
{
  "id": "6979d46501a3270436d37b98cf351439fbcbec8d5890d293dabfb8f85f723904",
  "question": "Cocher la (les) proposition(s) exacte(s) : Le métronidazole :",
  "answers": {
    "A": "Est un dérivé du pyrazole",
    "B": "Peut induire un effet antabuse",
    "C": "Peut être administré par voie parentérale intraveineuse",
    "D": "Peut être utilisé dans certaines parasitoses à protozoaires",
    "E": "Est inefficace dans les infections à germes anaérobies"
  },
  "correct_answers": ["B", "C", "D"],
  "nbr_correct_answers": 3,
}
```

# Métriques d'évaluation et système baseline

- Métriques d'évaluation
  - Tâche principale
    - Hamming score : prise en compte réponse incomplète (multi-labels)
    - Exact match ratio (EMR) : prise en compte réponse exacte
  - Tâche annexe
    - F1-macro et Accuracy
- Système baseline
  - CamemBERT
  - Classe majoritaire

# Tâche principale : Résultats et Classement

<b>Équipe</b>	<b>Hamming</b>	<b>Classement</b>	<b>EMR</b>	<b>Classement</b>
LIS	52,94	1	33,76	1
LIUM-IRISA	43,24	2	22,19	3
TTGV	41,54	3	23,95	2
SEQUOIA	37,93	4	15,27	4
ALMANACH-ARKHN	35,96	5	14,15	5
SPQR	24,93	6	9,97	6
Baseline	36,24	-	16,55	-
Classe majoritaire	23,93	-	13,67	-

# Tâche annexe : Résultats et Classement

<b>Équipe</b>	<b>F1-Macro</b>	<b>Classement</b>	<b>Accuracy</b>	<b>Classement</b>
LIS	42,42	1	68,65	1
TTGV	31,51	2	62,54	2
SPQR	22,99	3	47,43	3
Baseline	28,79	-	67,04	-
Classe majoritaire	13,62	-	51,61	-

# Tâches : Aucune restriction

<b>Équipe</b>	<b>Tâche principale</b>		<b>Tâche annexe</b>	
	<b>Hamming</b>	<b>EMR</b>	<b>F1-Macro</b>	<b>Accuracy</b>
LIS	85,17	72,83	71,57	79,58
SEQUOIA	64,40	46,46	44,36	65,92
Reproductible	52,94	33,76	42,42	68,65

# Programme

- **14h15 - 14h35** Equipe ALMANACH-ARKHN - Passe ta pharma d'abord ! *(Simon Meoni, Rian Touchent, Éric de la Clergerie)*
- **14h35 - 14h55** Equipe LIS : LIS@DEFT'23 : les LLMs peuvent-ils répondre à des QCM ? (a) oui; (b) non; (c) je ne sais pas. *(Benoit Favre)*
- **14h55 - 15h15** Equipe LIUM-IRISA : Qui de DrBERT, Wikipédia ou Flan-T5 s'y connaît le plus en questions médicales ? *(Clément Besnard, Mohamed Ettaleb, Christian Raymond, Nathalie Camelin)*
- **15h15 - 15h35** Equipe SEQUOIA : Participation d'EDF R&D au défi DEFT 2023 : réponses automatiques à des questionnaires à choix multiples à l'aide de « Larges Modèles de Langue » *(Meryl Bothua, Leila Hassani, Marie Jubault, Philippe Suignard)*
- **15h35 - 15h50** Pause café
- **15h50 - 16h10** Equipe SPQR : SPQR@Deft2013: Similarité Sorbonne Pour les Systèmes de Question Réponse *(Julien Bezançon, Toufik Boubehziz, Corina Chutaux, Oumaima Zine, Laurie Acensio, Andrea Briglia, Caroline Koudoro-Parfait, Gaël Lejeune)*
- **16h10 - 16h30** Equipe TTGV : Participation de l'équipe TTGV à DEFT 2023 : Réponse automatique à des QCM issus d'examens en pharmacie *(Andréa Blivet, Solène Degrutère, Barbara Gendron, Aurélien Renault, Cyrille Siouffi, Vanessa Gaudray-Bouju, Christophe Cerisara, Hélène Flamein, Gaël Guibon, Matthieu Labeau, Tom Rousseau)*
- **16h30 - 17h15** CLEF eHealth : 9 ans de benchmarks pour l'évaluation de la recherche d'information dans le domaine médical *(Lorraine Goeriot)*
- **17h15 - 17h30** Discussion et conclusion

# Conférencière invitée

- **Lorraine Goeuriot**

- MCF à l'Université Grenoble Alpes
- Laboratoire Informatique de Grenoble - Équipe MRIM
- ***CLEF eHealth : 9 ans de benchmarks pour l'évaluation de la recherche d'information dans le domaine médical***

