



**26**  
**AVR.**  
2023

🕒 de 12h à 13h

📍 LAENNEC - Amphi 1

## SÉMINAIRE

# Enrichissement linguistique et cognitif de modèles de langue contextualisés pour le domaine médical

Corentin Blanc

-

Grâce aux nouveaux progrès technologiques, des quantités vertigineuses de données textuelles sont créées chaque jour. Ces dernières pourraient offrir des opportunités immenses aux professionnels de santé, mais leur exploitation complexe et fastidieuse reste très peu répandue. Pour résoudre ce problème, les professionnels de santé se tournent donc peu à peu vers des solutions basées sur des modèles de langue contextualisés, la référence en matière d'outils de traitement automatique du langage naturel. Cependant, leur comportement dans le domaine médical demeure méconnu et insuffisamment étudié en langue française. L'objectif sera d'étudier le comportement de deux modèles français -FlauBERT et CamemBERT- puis de les enrichir linguistiquement et cognitivement afin de mieux répondre aux exigences spécifiques du domaine médical. Ces recherches ont abouti au développement de deux nouveaux modèles -BioFlauBERT et BioCamemBERT- qui ont une meilleure compréhension du langage naturel dans le domaine médical, ainsi qu'à une application industrielle prometteuse pour prédire des motifs de consultation.