



Intitulé de l'offre de stage

NLP pour l'exploitation des compte-rendus médicaux

| | |
|-------------------------------|---|
| Stagiaire en | IA (NLP) |
| Affectation | CHU Toulouse |
| Durée | 5 à 6 mois |
| Rémunération | environ 670 € net mensuel |
| Date de la publication | 15/10/2025 |
| Date d'embauche prévue | mars 2026 suivant disponibilités |
| Lieu | CHU, 2 rue Charles Viguerie, 31300 Toulouse, France |

Le Centre Hospitalier Universitaire de Toulouse

Le Centre Hospitalier Universitaire (CHU) de Toulouse est constitué de plusieurs sites (les principaux étant les sites de Rangueil, Larrey et Purpan, ainsi que celui de l'oncopole en commun avec l'institut Claudio Régaud); il comprend 4 000 médecins et 12 000 personnels hospitaliers. La mission de recherche et d'innovation du CHU fait partie intégrante à la fois de son activité quotidienne et de sa stratégie pour l'avenir. Elle est menée en collaboration étroite avec les facultés et les organismes de recherche que sont notamment l'Institut national de la santé et de la recherche médicale (Inserm) et le Centre national de la recherche scientifique (CNRS).

L'Institut de Recherche en Informatique de Toulouse

L'Institut de Recherche en Informatique de Toulouse (IRIT), une des plus imposantes Unité Mixte de Recherche au niveau national, est l'un des piliers de la recherche en Occitanie avec ses 700 membres, permanents et non-permanents. De par son caractère multi-tutelle (CNRS, Universités toulousaines), son impact scientifique et ses interactions avec les autres domaines, le laboratoire constitue une des forces structurantes du paysage de l'informatique et de ses applications dans le monde du numérique, tant au niveau régional que national.

L'équipe d'accueil

L'informatique, le traitement des données et l'intelligence artificielle sont appelés à prendre une place croissante dans le monde de la recherche médicale. Dans ce cadre, le CHU de Toulouse a créé le Centre de Données pour la Santé et la Recherche (CDSR) dont le but est le recueil, l'analyse et le traitement de la donnée médicale, en support aux activités de recherche médicales du CHU. Elle collabore activement avec l'IRIT afin d'appliquer les techniques les plus récentes de la science des données et de l'intelligence artificielle dans le domaine de la santé.

Le stagiaire intégrera l'équipe "Big Data" composée de 3 personnes (une data engineer, un data scientist, et un ingénieur devops), qui procède à la mise en qualité et à l'extraction des données, entre autres pour les finalités de recherche, et contribue à l'Entrepôt de Données de Santé du CHU de Toulouse.

Objet du stage

Les comptes-rendus médicaux créés au CHU de Toulouse sont majoritairement rédigés et stockés en texte libre. Dans le cadre de projets de recherche clinique, l'une des premières étapes, appelée "pré-screening", consiste à évaluer le volume de patients correspondant aux critères d'inclusion du projet afin d'en valider la pertinence et la faisabilité. Actuellement, uniquement des pré-screenings "simples" peuvent être réalisés, basés sur quelques données structurées, ou bien des expressions régulières sur le texte libre des comptes-rendus médicaux. Ainsi l'équipe Big Data souhaite développer un outil interne basé sur de l'intelligence artificielle permettant d'exploiter davantage le contenu en texte libre des comptes-rendus médicaux grâce à un LLM.

Le travail de l'étudiant consistera à participer à :

- > Faire un état de l'art des LLM
- > Constituer des jeux de documents pour du pré-screening
- > Faire un état de l'art, des expérimentations autour de l'indexation d'une base de documents et de son requêtage via LLM
- > Tester des LLM retenus sur les tâches de pré-screening
- > Comparer les performances

Perspectives et retombées attendues du projet :

- > Apporter la preuve de concept que l'identification des comptes-rendus pertinents via une requête textuelle est possible
- > Développement d'un outil de pré-screening intelligent

Formation

École d'ingénieur, de préférence avec spécialisation en informatique ou mathématiques (en année de césure ou stage long). Master 2 informatique ou mathématiques appliquées.

Compétences attendues

- > Maîtrise de langages de programmation (dont Python).
- > Connaissance des bases du Deep Learning.
- > Une expérience sur les librairies associées (Pytorch, transformers, etc.), une connaissance des bonnes pratiques de développement (ex : Git, conteneurisation, etc.), ainsi qu'une connaissance du NLP et de l'orchestration d'agents LLM seraient un plus.
- > Une expérience sur environnement ligne de commande Linux est recommandée.

Modalité de candidature

- > CV à envoyer à: nlp@stages-medecine-numerique.fr
- > Date limite de candidature : 01/01/2026
- > Encadrants: Laure Feuillet, Gaétan Faugère, Robin Schwob (CHU de Toulouse), Sandrine Mouysset (IRIT)

Une première sélection sera effectuée sur la base des CVs reçus. Les candidats seront immédiatement informés du résultat, et ceux qui seront retenus à l'issue de la première sélection seront invités à un ou deux entretiens individuels en téléconférence avant sélection définitive.

L'ensemble des offres de stage est disponible sur <http://www.stages-medecine-numerique.fr>.