



Intitulé de l'offre d'alternance

Exploitation des compte-rendus médicaux par NLP

Alternance en	Intelligence Artificielle
Affectation	CHU Toulouse
Durée	2 ans
Rémunération	suivant la réglementation en vigueur
Date de la publication	13/06/2026
Date d'embauche prévue	septembre 2026
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 Claudius 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 prennent une place très importante 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.

L'alternant sera également accueilli au sein du département d'information médical (DIM) du CHU de Toulouse. Le rôle du DIM est fondamental au sein de l'hôpital. Il est en charge du recueil et du codage de l'activité de l'hôpital ; à ce titre, il traduit les diagnostics et actes médicaux en codes standardisés (nomenclature CIM-10, CCAM), assure le financement de l'établissement dans le cadre de la Tarification à l'Activité, s'assure de la qualité des données, produit des indicateurs de performance et fournit des bases de données pour la recherche et les études cliniques.

Il sera encadré pour la partie "science des données" par Madame Sandrine Mouysset, chercheuse à l'IRIT en IA.

Objet du stage

Le DIM est, dans le cadre de ses missions, le principal exploitant des données médicales de l'établissement. Il a en particulier la mission de transformer les compte-rendus non structurés en codes structurés, et doit produire des indicateurs pour le pilotage interne et aussi pour la labellisation de l'hôpital dans le cadre de certifications internationales. A ce titre, il est le premier exploitant de techniques d'intelligence artificielle que sont les grands modèles de langage.

L'hôpital dispose de moyens de calcul puissants permettant de traiter en interne les dossiers médicaux en utilisant les très grands modèles de langages "Open Weights" comme gpt-oss-120b ou llama3. Le travail à réaliser sur deux ans consistera se composera de plusieurs parties :

- > Evaluer les différents modèles de langage OpenWeights en fonction des besoins précis de l'hôpital
- > Concevoir un pipeline permettant de réaliser un codage automatique CIM-10 des consultations ; aujourd'hui seuls les compte-rendus hospitaliers sont codés, mais le suivi des compte-rendus de consultation permettrait d'avoir une meilleure vision de l'activité hospitalière et permettrait également un meilleur suivi des pathologies chroniques

Formation

École d'ingénieur, de préférence avec spécialisation en intelligence artificielle, informatique ou mathématiques.

Compétences attendues

- > Maîtrise de langages de programmation (en particulier Python).

- > Connaissance de base en NLP
- > Une expérience sur les bibliothèques associées (Pytorch, API OpenAI, etc) une connaissance des bonnes pratiques de développement (ex : Git, conteneurisation, etc.),
- > Une expérience sur environnement ligne de commande Linux est recommandée.

Modalité de candidature

- > CV à envoyer à : dim@stages-medecine-numerique.fr
- > Date limite de candidature : 01/07/2026
- > Encadrants : Professeur Laurent Molinier (CHU de Toulouse), Jean-Marc Alliot (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>.