

Type d'offre : Offre en laboratoire

Date de publication : 19.11.25

IJCLab

Offre de stage - Development of software for an astronomical application and preparation of the corpus for an AI Large Language Models

Informations générales

Type de contrat : Stage

Durée du contrat : 5-6 mois

Niveau d'étude : M1/M2

Contact :

[S. Antier](#) / [C. Douzet](#)

Date de prise de poste : mer 01/07/2026 - 12:00

Métier : Technicien

Thématique : Supervision, contrôle, optimisation

IJCLab :

Le [Laboratoire de Physique des 2 Infinis Irène Joliot-Curie](#), ou IJCLab, est une Unité Mixte de Recherche du CNRS, de l'Université Paris Saclay et de l'Université Paris-Cité, située sur le campus de la Faculté des Sciences d'Orsay. Ce laboratoire est issu de la fusion de de cinq laboratoires (CSNSM, IMNC, IPNO, LAL et LPT) géographiquement et thématiquement proches les uns des autres sur le campus d'Orsay. Ces laboratoires partagent une histoire commune, liée à la création puis au développement du Campus d'Orsay.

IJCLab, qui rassemble environ 730 personnes, couvre les activités présentes auparavant dans ces cinq laboratoires. L'identité d'IJCLab est centrée sur le domaine de « la physique des deux infinis » et de leurs applications, avec toute la richesse des thématiques qui constitue cette physique. Cela se traduit par la présence de pôles historiques forts, de pôles liés à des thématiques émergentes et d'activités aux interfaces. Ce laboratoire a la capacité, la vocation et l'ambition d'avoir un impact mondial fort sur un large éventail de domaines scientifiques et techniques, en étant le moteur de grands projets phares au niveau national et international. Il encourage et aide également au soutien de projets à des échelles plus locales et des cycles plus rapides, susceptibles d'apparaître en fonction de l'actualité scientifique et/ou des innovations techniques.

Détail de l'offre (poste, mission, profil) :

Contexte de l'offre

De nos jours, les astronomes capturent une multitude de phénomènes transitoires tels que des étoiles dévorant leurs compagnons, des objets capturés par des trous

noirs supermassifs, des étoiles mourantes ou encore des collisions d'étoiles mortes. Parmi ceux-ci, les phénomènes à haute énergie peuvent illuminer le ciel pendant plusieurs heures à plusieurs jours. Ces événements rares sont très minoritaires mais d'une valeur inestimable pour plusieurs domaines scientifiques, comme l'étude de l'origine des éléments lourds présents sur Terre ou celle de l'Énergie Noire. C'est pourquoi nous devons collecter des ensembles de données grâce à une coordination entre l'espace et le sol.

Pour y parvenir, nous utilisons une application web nommée **SkyPortal** (<https://skyportal.io/>), utilisée à la fois aux États-Unis et en Europe.

Dans ce stage, dans le cadre du projet interdisciplinaire **MAFORAI** (Monitoring Astronomical Follow-up Of Rare events with AI) et en utilisant SkyPortal (une plateforme de suivi et de data science pour l'astronomie du domaine temporel), nous visons à relever le défi de la coordination du suivi en astronomie à l'aide d'un pipeline basé sur l'IA et alimenté par un grand modèle de langage (LLM).

L'objectif à long terme est de fournir une assistance automatisée aux astronomes en utilisant les informations collectées dans SkyPortal pour suggérer des stratégies d'observation, mais aussi pour mieux analyser les images recueillies au cours de ces campagnes.

Pour ce stage, le travail portera sur la première étape du projet : la création d'un premier corpus et de l'infrastructure de données nécessaire pour construire ce corpus. Cela inclut l'architecture permettant d'extraire, d'organiser et d'annoter les informations issues de campagnes d'observation passées trouvées dans SkyPortal. Ces produits de données sont actuellement dispersés dans de multiples sources, formats et modalités, incluant des types de données hétérogènes tels que du texte, des images, des journaux de commandes, des communications (comme des mises à jour instantanées de mission ou des rapports informels). Un défi majeur consiste à traiter cette hétérogénéité et à garantir que l'information extraite puisse être rendue reproductible pour entraîner un LLM.

Selon les intérêts du ou de la stagiaire, le corpus pourra se concentrer plus spécifiquement sur :

- des données textuelles, comme des conversations, journaux de décisions, avec une attention particulière portée à l'évolution temporelle des données de la campagne ;
- de la vision, comme des images combinées à des commentaires d'experts / métadonnées.

Nous demandons une expérience préalable en **Python**, **GitHub** et **API de LLM** (par exemple OpenAI).

Dans l'ensemble, ce stage offre l'opportunité de définir les fondations d'un système de coordination de suivi assisté par IA, avec un accent sur la création de corpus, l'organisation des données et l'architecture système en phase initiale.

URL de l'offre :

<https://www.dataia.eu/sites/default/files/2025-11/Internship%20-%20MAFORAI%20-%20Skyportal-1.pdf>

Lien vers l'offre sur le site dataia.eu : <https://da-cor-dev.peppercube.org/node/1462>