

L'Université PSL recrute 1 ingénieur de recherche en science des données pour les sciences sociales

Dans le cadre du développement de l'Institut ACSS, l'Université PSL recrute un ingénieur de recherche (IR) en science des données. Elle/il sera chargé(e) de mettre en œuvre des méthodes et outils de collecte et traitement de données issues de sources variées (Web, bases de données institutionnelles, archives, etc.). Elle/il aura également la responsabilité de veiller au respect des bonnes pratiques en matière de développement et de gestion du code et des données. Enfin, elle/il contribuera au développement de modèles statistiques ou d'apprentissage automatique (notamment dans le domaine du traitement automatisé de la langue naturelle).

Créé au sein de l'Université Paris Sciences et Lettres (PSL) et hébergé à Paris Dauphine, l'Institut « Applied Computational Social Sciences » a pour vocation de renforcer la recherche sur les grandes problématiques sociétales (cohésion politique et sociale, transition écologique, transformation numérique, efficacité et compétitivité économique) en articulant sciences des données et sciences sociales.

L'Institut collecte et traite à large échelle des données hétérogènes tant pour permettre des avancées scientifiques que pour contribuer à éclairer le débat public et la décision. Il réunit sur une équipe pluridisciplinaire de chercheurs en sciences sociales et s'appuie sur une équipe d'ingénieurs en sciences de données qui apportent leurs expertises pour constituer des bases de données originales et opérer des traitements complexes. Ces projets sont initiés et portés par des laboratoires du CNRS, de Dauphine, de l'ENS, de l'INSP et des MinesParis-Tech. Les résultats des travaux ont vocation à être largement diffusés auprès de partenaires institutionnels et du monde économique.

Procédure de candidature :

Le formulaire de candidature est disponible à cette adresse : <https://acss-dig.psl.eu/fr/candidate>. Il devra inclure un CV, d'une lettre de motivation et du relevé des derniers diplômes. L'examen des candidatures commencera le **2 novembre 2024** et se poursuivra jusqu'à ce que les postes soient pourvus.

Prise de fonctions le 1^{er} janvier (ou février) 2025.

Contact : Bruno Chaves Ferreira bruno.chavesferreira@dauphine.psl.eu

Formation et compétences requises :

Compétences nécessaires :

- Expérience dans le développement de réseaux de neurones profonds et autres modèles statiques avancés appliqués au traitement automatisé du langage sur de larges corpus.
- Maîtrise des écosystèmes Python et/ou R dédiés à la science des données.
- Plus spécifiquement en Python, maîtrise de numpy, pandas, pytorch et l'environnement hugging face.
- En R, maîtrise du tidyverse, de tidymodels et des bibliothèques associées, ainsi que de torch.
- Maîtrise des bases de données relationnelles et NoSQL.
- Compréhension des méthodes scientifiques des sciences humaines et sociales.

Compétences additionnelles souhaitées :

- Expérience dans le réglage des hyperparamètres pour les grands modèles de langage (LLM) afin d'améliorer les performances et l'efficacité.
- Familiarité avec le « fine-tuning » de modèles basés sur des architectures de type « Transformers ».
- Connaissance des techniques d'entraînement distribué et de l'exploitation des GPU dans un cluster.
- Pratique des services web modernes (architecture REST, formats JSON, XML, etc.).
- Méthodologies de développement : contrôle de version (GIT), tests unitaires, intégration continue, etc.

Formation :

Diplôme d'ingénieur en informatique (ou mathématiques/statistiques) avec au moins 3 ans d'expérience ou titulaire d'un doctorat avec une thèse dans le domaine.

Divers :

Il s'agit d'une création de poste sur CDD pour une durée de 1 an renouvelable. Rémunération suivant profil et expérience en fonction des grilles CNRS. Le poste est basé à l'université de Paris Dauphine.